

**Ghid de calitate al statisticii
oficiale din România**

*Quality Guidelines for Romanian
Official Statistics*

Autori / colaboratori

Authors / collaborators

Din partea INS

From INS

Gabriela POPA
Răzvan Andrei PAVEL
Claudia CĂPĂȚINĂ
Iuliana Ramona GRIGORE
Maria RĂDULESCU
Virginia BALEA
Andreea CAMBIR
Nina ALEXEVICI
Florentina GHEORGHE

Din partea consorțiului

From the consortium

Michael John COLLEDGE (ICON)
Manfred EHLING (ICON)

Revizuire editorială

Editorial revision

Andreea DRĂGUȘIN (ICON)
David GRAU PEÑAS (ICON)

Beneficiar

Institutul Național de Statistică
Direcția generală de organizare și coordonare
a sistemului statistic național

Beneficiary

*National Institute of Statistics
Directorate General of Organisation and
Coordination of the National Statistical System*

Prefață

Misiunea Institutului Național de Statistică este de a satisface nevoile de date statistice ale utilizatorilor de toate tipurile, prin culegerea, prelucrarea și diseminarea datelor relevante, precise, coerente, oportune și accesibile. În acest context, este vital ca managementul calității datelor să fie realizat printr-o metodologie solidă.

Scopul acestui document este de a uni o gamă largă de bune practici, în cadrul unui ghid, care să poată veni în sprijinul managerilor de program și al personalului acestora, în gestionarea calității și care să poată asigura dovezi ale culturii calității în cadrul Institutului Național de Statistică. Documentul este conceput pe baza unor materiale furnizate de Sistemul Statistic European și de alte ghiduri naționale privind calitatea.

Ghidul trebuie aplicat în mod selectiv. Acesta este mai degrabă complementar, decât un înlocuitor al expertizei și opiniilor personalului din Institutul Național de Statistică. Proiectarea și implementarea programelor cercetărilor statistice necesită o echilibrare atentă a obiectivelor (adesea contradictorii) privind relevanța, acuratețea, oportunitatea, accesibilitatea, claritatea, coerența, în timp ce se încearcă minimizarea costurilor și a sarcinii de răspuns.

Doresc să mulțumesc experților care au contribuit la elaborarea documentului și în mod special celor de la Institutul de Statistică al Canadei, pentru că ne-au pus la dispoziție Ghidul Calității, ca fundament pe care s-a construit prezentul document.

Preface

The mission of the National Institute of Statistics is to satisfy the statistical data needs of users of all types through the collection, processing and dissemination of relevant, accurate, coherent, timely and accessible data. In this context management of data quality through sound survey methodology is vital.

The aim of this document is to bring together a wide range of good practices into guidelines that can assist program managers and their staff in managing quality and that provide evidence of the quality culture at the National Institute of Statistics. The document draws on material from the European Statistical System and from other national quality guidelines.

The guidelines have to be selectively applied. They are complementary to rather than a replacement for expertise and judgement of National Institute of Statistics staff. Designing and implementing survey programs requires careful balancing of (often competing) objectives of relevance, accuracy, timeliness, accessibility, clarity, coherence, whilst at the same time attempting to minimise cost and reporting burden.

I would like to thank all the experts who have contributed to the document and especially Statistics Canada for permission to use its Quality Guidelines as the foundation upon which the document is based.



Prof. univ. dr. Vergil VOINEAGU

Președintele Institutului Național de Statistică
President of the National Institute of Statistics

Considerații

Numeroase persoane și-au adus contribuția la organizarea și succesul proiectului "Dezvoltarea continuă a sistemului statistic românesc" pe planul îmbunătățirii managementului calității statisticii românești. Instituția beneficiară a fost Institutul Național de Statistică (INS), responsabil pentru colaborarea zilnică cu experții internaționali și pentru furnizarea documentelor metodologice și a datelor necesare pentru îndeplinirea sarcinilor proiectului. La proiect au participat în mod special următoarele persoane din INS: d-na Gabriela Popa (coordonator de echipă), d-l Răzvan Pavel, d-na Claudia Căpățînă și d-na Iuliana Ramona Grigore, sub coordonarea d-nei Sândica Neagu.

Contractorul proiectului, **Consortiul ICON**, a acordat asistența tehnică și a asigurat implementarea generală a proiectului. Consortiul a fost reprezentat de ICON-INSTITUTE Public Sector GmbH (ca lider), și Centrul Român de Modelare Economică (CERME). Expertiza tehnică privind ghidul calității a fost asigurată de d-l Michael Colledge, autorul principal al acestui document, și de d-l Manfred Ehling, expert principal al acestei componente a proiectului. Echipa de consultanți a fost condusă de Dr Wolfgang Kadgien (Managerul proiectului) și d-l José Cervera-Ferri (șeful de echipă).

Punctul de plecare pentru elaborarea acestui document l-a constituit *Ghidul Calității* al Statisticii Canadiene, a patra ediție 2003, *Catalog nr.12-539-XIE*, utilizat cu permisiunea Statisticii Canadiene. Au fost adăugate informații suplimentare și s-au efectuat modificări semnificative pentru a adapta documentul la situația din România. Așadar, acest document nu este prezentat ca fiind conceput prin afiliere cu Statistica din Canada, iar Statistica din Canada nu este implicată în niciun fel în documentul rezultat. Documentul original aparținând instituției Canadiene poate fi vizualizat pe website-ul Statisticii Canadiene, la adresa www.statcan.gc.ca.

Acknowledgements

Many persons have contributed in the organization and success of the project "Continuous Development of the Romanian Statistical System" in relation to improving quality management of Romanian Statistics. The beneficiary body has been the National Institute of Statistics (NIS), responsible for the day-to-day collaboration with the consultants and for providing methodological documents and data necessary to perform the tasks of the project. In particular, the following NIS staff has participated: Ms Gabriela Popa (team coordinator), Mr Răzvan Pavel, Ms Claudia Căpățînă and Ms Iuliana Ramona Grigore, under the supervision of Ms Sândica Neagu.

The contractor of the project, **the ICON-Consortium**, has provided technical assistance and overall implementation of the project. The consortium was composed by ICON-INSTITUTE Public Sector GmbH (as leader), and the Romanian Centre for Economic Modelling (CERME). Technical expertise on quality guidelines was provided by Mr Michael Colledge, who is the main author of this document, and Mr Manfred Ehling, key expert on the project component. The team of consultants was led by Dr Wolfgang Kadgien (Project Manager) and Mr José Cervera-Ferri (Team leader).

The starting point for production of this document was Statistics Canada's *Quality Guidelines, Fourth Edition 2003, Catalogue no. 12-539-XIE*, which was used with the permission of Statistics Canada. Much material has been added and significant changes made to tailor the document to the Romanian situation. Thus this document is not presented as having been made in affiliation with Statistics Canada, nor does Statistics Canada in any way endorse the result. The original Statistics Canada document may be seen on Statistics Canada's website at www.statcan.gc.ca.

Cuprins

Prefață	1
Considerații.....	2
Abrevieri și acronime	5
1 Introducere	7
2 Conceptele calității	9
2.1 Perspectiva internațională.....	9
2.2 Introducerea managementului calității în INS	10
2.3 Politica privind calitatea	11
2.4 Modelul calității.....	12
2.5 Evaluarea și raportarea calității.....	14
3 Ghidul calității.....	16
3.1 Obiective, utilizări și utilizatori.....	16
3.2 Concepte, variabile și clasificări	18
3.3 Acoperire și cadrul de eșantionare.....	19
3.4 Eșantionarea	21
3.5 Proiectarea chestionarelor	24
3.6 Culegerea și captarea de date.....	26
3.7 Non-răspunsul	28
3.8 Editarea.....	30
3.9 Imputarea	31
3.10 Estimarea	33
3.11 Ajustarea sezonieră și estimarea ciclului de trend.....	35
3.12 Evaluarea calității datelor.....	38
3.13 Controlul confidențialității	40
3.14 Analiza și prezentarea datelor.....	43
3.15 Diseminarea datelor	45
3.16 Documentarea	46
3.17 Utilizarea datelor administrative.....	48
4 Cadrul General de Management al Calității.....	51
4.1 Managementul Relevanței	51
4.2 Managementul Acurateței	52
4.3 Managementul Oportunității și Punctualității	53
4.4 Managementul Accesibilității	54
4.5 Managementul Clarității	54
4.6 Managementul Coerenței și Comparabilității.....	55
4.7 Managementul Raportării Calității	56
4.8 Managementul Cadrului Instituțional.....	57
5 Anexă: Politici privind Calitatea Sistemului Statistic European	58
5.1 Declarația privind calitatea SSE	58
5.2 Codul de Practici al Statisticilor Europene.....	59

Table of contents

Preface	1
Acknowledgements.....	2
Abbreviations and acronyms	5
1 Introduction.....	67
2 Quality Concepts.....	69
2.1 International Perspective	69
2.2 Introducing Quality Management at the NIS.....	70
2.3 Quality Policy.....	71
2.4 Quality Model.....	72
2.5 Quality Assessment and Reporting	74
3 Quality Guidelines	75
3.1 Objectives, Uses and Users.....	75
3.2 Concepts, Variables and Classifications	77
3.3 Coverage and Frames.....	78
3.4 Sampling.....	80
3.5 Questionnaire Design.....	83
3.6 Data Collection and Capture.....	85
3.7 Non-Response.....	86
3.8 Editing.....	88
3.9 Imputation.....	90
3.10 Estimation.....	91
3.11 Seasonal Adjustment and Trend-Cycle Estimation	94
3.12 Data Quality Evaluation	96
3.13 Disclosure Control	98
3.14 Data Analysis and Presentation	101
3.15 Data Dissemination	103
3.16 Documentation	104
3.17 Administrative Data Use.....	105
4 Quality Management Framework.....	109
4.1 Managing Relevance	109
4.2 Managing Accuracy.....	110
4.3 Managing Timeliness and Punctuality	111
4.4 Managing Accessibility	112
4.5 Managing Clarity.....	112
4.6 Managing Coherence and Comparability.....	113
4.7 Managing Quality Reporting	113
4.8 Managing the Institutional Environment.....	115
5 Appendix: European Statistical System Quality Policies	116
5.1 ESS Quality Declaration	116
5.2 European Statistics Code of Practice	117

Abrevieri și acronime

Abbreviations and acronyms

ESS Code of Practice	CoP	<i>Cod de Practici al Statisticilor Europene</i>
ESS Checklist for Survey Managers	DESAP	<i>Lista de control SSE pentru managerii de cercetări statistice</i>
Computer Assisted Personal Interviewing	CAPI	<i>Interviu asistat de calculator</i>
Computer-assisted telephone interviewing	CATI	<i>Interviu telefonic asistat de calculator</i>
Electronic data reporting	EDR	<i>Raportare electronică a datelor</i>
European Foundation for Quality Management	EFQM	<i>Fundația Europeană pentru Managementul Calității</i>
ESS Handbook for Quality Reports	EHQR	<i>Manual SSE privind rapoartele de calitate</i>
Euro SDMX Metadata Structure	ESMS	<i>Structura de metadata Euro SDMX</i>
ESS Standard Quality Indicators	ESQI	<i>Indicatorii standard de calitate ai SSE</i>
ESS Standard for Quality Reports	ESQR	<i>Standardul SSE privind rapoartele de calitate</i>
European Statistical System	ESS	<i>Sistemul Statistic European</i>
Intelligent character recognition	ICR	<i>Recunoașterea inteligentă a caracterelor</i>
Metadata Common Vocabulary	MCV	<i>Vocabularul comun de metadata</i>
National Institute of Statistics	NIS	<i>Institutul Național de Statistică</i>
Quality Management Framework	QMF	<i>Cadrul general de management al calității</i>
Statistical Data and Metadata Exchange	SDMX	<i>Sistemul de schimb de date și metadata statistice</i>
Total quality management	TQM	<i>Managementul Calității Totale</i>

1 Introducere

Scopul Ghidului

Prezentul document reunește mai multe aspecte ce trebuie luate în considerare atunci când se urmăresc obiectivele calității în Institutul Național de Statistică (INS). Acesta se axează pe modul în care trebuie asigurată calitatea printr-o proiectare și o implementare eficientă și adecvată ale cercetărilor și ale altor procese statistice, de-a lungul întregului lor ciclu de viață. Obiectivele specifice sunt formularea conceptelor calității și promovarea conștientizării calității, asigurarea îndrumării tehnice privind proiectarea și realizarea cercetărilor statistice și sintetizarea cadrului de asigurare a calității în organizație.

În acest context termenii *cercetare statistică* și *proces statistic* se referă la cercetarea statistică prin sondaj, recensământ, cercetare statistică bazată pe date administrative, producerea indicilor de preț sau a altor indici economici, sau orice altă lucrare statistică de sinteză, precum Conturile Naționale și Balanța de Plăți, așa cum se va discuta în cele ce urmează.

Utilizatori și utilizări

Ghidul este în primul rând destinat utilizării de către personalul angajat în proiectarea și/sau realizarea cercetărilor statistice prin sondaj, recensămintelor și altor procese statistice. El poate, de asemenea, servi managerilor de la cel mai înalt nivel și utilizatorilor care doresc să se asigure că metodele de cercetare sunt adecvate, iar calitatea rezultatului este suficient de bună.

Terminologie

În măsura în care este posibil, definițiile termenilor sunt în conformitate cu Glosarul Sistemului Statistic European (SSE). Atunci când un termen nu se află în Glosarul SSE, definiția sa este extrasă dintr-o altă sursă internațională disponibilă, de exemplu Vocabularul Comun de Metadate (VCM), creat de Organizația pentru Cooperare și Dezvoltare Economică (OECD) și alte organizații internaționale care se ocupă de inițiativele privind Schimbul de Date și Metadate Statistice (SDMX). Glosarul statistic al INS este luat de asemenea în considerare.

Tipuri de cercetări statistice

Din perspectiva tehnică, metodele INS de producere a statisticilor prezintă o mare diversitate. În consecință, nu se poate aplica tuturor proceselor de producție un singur set de recomandări, în special acelea privitoare la acuratețe, ci este necesară introducerea unor diferențieri. În scopul elaborării acestui document și pe baza *Manualului SSE privind rapoartele de calitate* (Eurostat, 2009), se disting șase tipuri de cercetări și procese statistice, după cum urmează:

1. *Cercetarea statistică prin sondaj*. Aceasta este o cercetare care se bazează (de obicei) pe o procedură de eșantionare probabilistică, implicând culegerea directă a datelor de la respondenți. Pentru acest tip de cercetare există o teorie atestată privind acuratețea, prin care componentele erorii sunt bine definite (erori de eșantionare și de non-eșantionare). Ca exemple, menționăm Ancheta asupra Forței de Muncă în Gospodării, Ancheta Costului Forței de Muncă și Ancheta Investițiilor Străine Directe.
2. *Recensământul*. Acesta poate fi privit ca un caz special al cercetării prin sondaj, în care datele sunt colectate de la toate unitățile din cadrul populației țintă. Un exemplu este Recensământul Populației și Locuințelor din 2002.
3. *Cercetarea statistică ce utilizează sursele administrative*. Acest tip de cercetare utilizează datele colectate pentru alte scopuri decât producerea directă de statistici.
4. *Cercetări statistice ce implică surse multiple de date*.
5. *Indicii de preț sau alți indici economici privind procesul de producție*.

6. *Lucrări statistice de sinteză.* Datele sunt colectate dintr-o varietate de surse primare, pentru a produce date agregate cu o semnificație conceptuală specială, de exemplu Conturile Naționale și Balanța de Plăți.

Conținutul Ghidului

Capitolul 2 urmărește plasarea Ghidului într-un context internațional. Acesta descrie conceptele de bază ale calității, pe care se bazează abordarea INS privind calitatea, incluzând o trecere în revistă a documentației SSE și a altor documentații internaționale disponibile privind calitatea și definițiile calității rezultatelor și calității proceselor. Acesta este în mare parte extras din *Manualul SSE privind Rapoartele de Calitate* (Eurostat, 2009).

Capitolul 3 prezintă principiile tehnice de utilizare în contextul proiectării, reproiectării, dezvoltării sau punerii în practică a unei cercetări sau a altor procese statistice. Ghidul este elaborat în principal pentru recensăminte și cercetări statistice prin sondaj, însă se aplică și altor tipuri de cercetări statistice dintre cele definite anterior.

În timp ce Capitolul 3 se axează pe cercetările individuale și pe alte procese statistice, Capitolul 4 se referă la întreaga organizație, descriind cadrul calității acesteia.

Anexele conțin copii a două documente relevante privind politica SSE.

2 Conceptele calității

2.1 Perspectiva internațională

Evoluția Managementului Calității

Managementul modern al calității își are originea în *controlul statistic al calității* proiectat pentru liniile de producție din fabricile anilor '20 și care încă funcționează. Walter Shewhart, cunoscut uneori ca "părinte al controlului statistic al calității", introduce noțiunea de variație datorată cauzelor bine determinate și celor aleatoare și ulterior eliminarea cauzelor bine determinate, folosind diagramele de control ale calității procesului. Esențiale în controlul calității sunt inspecția și corecția. Ulterior, între anii 1940 și 1950, s-a acordat o mai mare însemnătate prevenirii defectelor apărute decât corectării acestora, de pildă utilizarea feed-back-ului de la rezultatele obținute în urma inspecției calității, ca bază pentru modificarea problemelor apărute la liniile de producție și pentru instruirea angajatului. Această procedură a fost denumită *controlul calității în amonte*, iar gama mai largă de măsurare a calității a fost denumită *asigurarea calității*.

În următorii 30 de ani, ca rezultat al lucrărilor scrise și al sfaturilor date de către „guru” ai calității, precum William Edwards Deming, recunoscut pentru lucrarea sa *Paisprezece principii de management* și Joseph Juran, care a militat pentru dimensiunea umană, noțiunea de asigurare a calității a fost extinsă mai departe la *Managementul Calității Totale (TQM)*. Principiile TQM sunt prezentate în cele ce urmează:

- *Focalizarea pe client*: o organizație depinde de clienții săi și astfel trebuie să înțeleagă și să se străduiască să satisfacă nevoile acestora; clienții sunt esențiali în determinarea a ceea ce constituie o bună calitate; calitatea este ceea ce percepe clienții, mai degrabă decât organizația.
- *Leadership și consistența scopurilor*: liderii stabilesc unitatea scopului și direcției unei organizații; aceștia trebuie să creeze și să mențină un mediu intern, care să permită personalului să se implice total în atingerea obiectivelor organizației; îmbunătățirea calității necesită o bună conducere și o direcție susținută.
- *Implicarea oamenilor*: oamenii de la toate nivelurile sunt esența unei organizații; implicarea lor totală permite ca abilitățile lor să fie total valorificate.
- *Orientarea către proces*: gestionarea activităților și a resurselor ca proces este eficientă; orice proces se poate diviza într-un lanț de sub-procese, pentru care rezultatul unui proces este input pentru următorul.
- *Abordarea sistematică către management*: identificarea, înțelegerea și gestionarea proceselor ca sistem contribuie la operativitate și eficacitate.
- *Îmbunătățire continuă*: îmbunătățirea continuă ar trebui să fie un obiectiv permanent al unei organizații.
- *Abordarea bazată pe fapte în luarea deciziilor*: deciziile eficace se bazează pe analiza informațiilor și datelor.
- *Relațiile reciproc avantajoase cu furnizorul*: o organizație și furnizorii săi sunt interdependenți, iar o relație cu avantaj reciproc contribuie la dezvoltarea amândurora.

Seriile ISO 9000 ale sistemului de management al calității

Specificația principiilor TQM dată mai sus a fost extrasă din standardul ISO 9000, standardul de calitate cel mai larg utilizat în lume. Sunt de fapt trei standarde:

- *ISO 9000: 2005 Sistemele de management al calității – Elemente esențiale și Vocabular*: descrie elementele fundamentale ale sistemelor de management al calității, incluzând principiile calității (așa cum sunt reproduse mai sus), terminologie și modele;

- *ISO 9001: 2000 Sistemele de management al calității – Cerințe:* specifică caracteristicile cerute pentru un sistem de management al calității. Este cel mai bine cunoscut dintre toate cele trei standarde, întrucât este cel vizat de către organizații pentru obținerea certificării;
- *ISO 9004: 2000 Sistemele de management al calității – Principii:* este proiectat să asiste organizațiile care au deja implementate sisteme de management al calității, pentru a obține îmbunătățiri în ceea ce privește îmbunătățirea performanței.

Standardul EFQM al calității

Fundația Europeană pentru Managementul Calității a produs ceea ce este de fapt un standard al calității întrucâtva similar standardului ISO 9000. Cu toate că inițial a fost proiectat pentru utilizarea de către organizațiile comerciale din Uniunea Europeană, *Modelul de Excelență al EFQM* este în prezent recunoscut la nivel mondial, mii de organizații din toată lumea utilizând principiile EFQM pentru a-și ghida strategiile și operațiunile afacerilor lor. *Excelența* este definită ca *practica remarcabilă în gestionarea organizației și obținerea de rezultate* și se spune că ar necesita angajamentul conducerii în respectarea elementelor principale în mod similar cu principiile ISO 9000, respectiv: orientarea către rezultate; focalizarea pe client; conducerea și consistența scopului; managementul orientat pe procese și fapte; dezvoltarea și implicarea oamenilor; învățarea continuă, inovarea și îmbunătățirea; dezvoltarea parteneriatului; responsabilitatea socială corporatistă.

ISO 20252:2006 Piața, opinia și cercetarea socială – Vocabular și cerințe

Noul standard internațional stabilește terminologia și cerințele pentru organizațiile ce realizează cercetări de piață, sondaje de opinie și cercetări sociale și stabilește un nivel comun al calității pentru cercetarea de piață la nivel global. Își are originile în standardul BS7911 al Institutului Britanic de Standardizare, pe care îl înlocuiește. Cuprinde secțiuni privind sistemele de management al calității, gestionarea cercetării, culegerea de date (inclusiv munca de teren, auto-completarea chestionarelor și culegerea din surse secundare), managementul, prelucrarea și raportarea datelor (captarea, codificarea, editarea, analizarea, securizarea).

2.2 Introducerea managementului calității în INS

Nevoia de interpretare a standardelor internaționale de calitate

Cu excepția standardului ISO 20252, standardele și modelele de calitate descrise în secțiunea anterioară sunt destinate acoperirii tuturor organizațiilor, indiferent de structura lor organizațională, de procesele și produsele lor și sunt exprimate în termeni foarte generali. Astfel, așa cum se precizează în introducerea la ISO 9001:2000:

- Proiectarea și implementarea de către o organizație a sistemului de management al calității sunt influențate de diverse nevoi, obiective specifice, de produsele furnizate, procesele angajate și de dimensiunea și structura organizației. Nu este intenția standardului de a implica uniformitate în structura sistemelor de management al calității sau uniformitate în documentație.

Astfel, în dezvoltarea sistemului său de management al calității (denumit *cadru general de management al calității (QMF)*), INS trebuie să interpreteze standardele internaționale în contextul său specific.

Contextul INS

Contextul în interiorul căruia operează INS, poate fi caracterizat după cum urmează:

- Este o instituție de stat, nu privată și care nu se bazează pe profit. Furnizează mai degrabă datele gratuit *utilizatorilor*, decât contra cost. Astfel, utilizatorii nu pot influența calitatea datelor prin decizii de achiziționare.

- Unii utilizatori sunt de fapt *utilizatori interni*, de exemplu domeniul conturilor naționale este un utilizator, dar și un producător.
- Principalele intrări la nivel național sunt, de obicei, date de la întreprinderile individuale, gospodării și persoane, fie colectate direct, fie prin intermediul proceselor administrative.
- Procesele cheie de producție reprezintă conversia datelor individuale în date agregate, adică statisticile și gruparea lor în rezultate statistice naționale.
- Produsele primare, denumite în prezentul document *rezultate*, sunt statisticile și serviciile anexe.

Elemente ale Cadrului General al Managementului Calității în INS

Luând în considerare aceste caracteristici, cadrul INS de management al calității (QMF) ar trebui să cuprindă:

- Politica privind calitatea – o scurtă declarație din partea managementului de top, care să indice nivelul angajamentului către managementul calității;
- Modelul calității – o definiție a ceea ce se înțelege prin calitate, elaborată de obicei în termenii componentelor calității;
- Obiectivele calității, standardele și ghidul calității – obiectivele țintă ale calității împreună cu standardele internaționale sau locale și ghidul adoptat de INS;
- Procedurile de asigurare a calității – parte a/încorporate în procesele de producție în cea mai mare măsură;
- Procedurile de evaluare a calității – uneori încorporate în procedurile de asigurare a calității, realizate cu regularitate, de exemplu pe baza unei liste de control pentru auto-evaluare, cum ar lista DESAP;
- Procedurile de măsurare a calității – în mod specific incluzând un set de indicatori de calitate și performanță, cu proceduri de culegere a valorilor datelor corespunzătoare, fiind încorporate în procesele de producție în cea mai mare măsură;
- Programul de îmbunătățire a calității – inițiativele de îmbunătățire continuă și re tehnologizare.

Scopul este de a ordona toate aceste elemente, așa cum vor fi discutate în următoarele paragrafe și în Capitolul 4.

2.3 Politica privind calitatea

Politici internaționale

Comisia de Statistică a Națiunilor Unite asigură consiliere pe tema cadrului instituțional în interiorul căreia se poate dezvolta managementul calității prin *Principiile fundamentale ale Națiunilor Unite ale Statisticilor Oficiale* pentru sistemele statistice naționale, promulgate în 1994. Există zece principii.

În SSE, *Declarația de calitate a SSE* a fost adoptată de Comitetul pentru Programul Statistic în 2001 ca măsură oficială către managementul calității totale în SSE, în conformitate cu Modelul de Excelență EFQM. Conține doar o pagină și jumătate, cuprinzând declarația misiunii și a viziunii SSE și zece principii împrumutate de la Principiile Fundamentale ale Națiunilor Unite și ajustate special pentru contextul SSE.

Codul de Practici al Statisticilor Europene (CoP) a fost promulgat de către Comisia Europeană în 2005. Acesta obligă Eurostat și cele 27 de State Membre, la respectarea celor cincisprezece principii ce acoperă *mediul instituțional, procesele și rezultatele statistice*.

Calitatea este de asemenea, un considerent în cadrul altor reglementări adoptate de Consiliul și Parlamentul European, creând baza legală pentru realizarea statisticilor europene în diverse domenii. Reglementările Consiliului sunt ele însele mecanisme de asigurare a calității, stabilind norme specifice, de actualitate, prin crearea unor standarde metodologice, ce conduc la creșterea acurateței și comparabilității, precum și acoperirea relevanței conform nevoilor instituțiilor europene privind statisticile naționale.

Politica privind calitatea în INS

În calitate de coordonator al Sistemului Statistic Național românesc, INS are misiunea de a satisface în cea mai mare măsură nevoia de informații pentru toate categoriile de utilizatori de date statistice prin culegerea, producerea și diseminarea datelor statistice relevante, precise, coerente, oportune și accesibile, necesare luării de decizii privind dezvoltarea economică și socială a țării și privind realitățile societății românești.

În ultimii șase ani, INS a înregistrat un progres considerabil în direcția managementului calității totale și asigurării unei culturi a calității în interiorul organizației. Ținta în următorii ani este de a obține performanțe metodologice și operaționale la niveluri comparabile cu cele mai avansate institute naționale de statistică din Statele Membre ale UE.

2.4 Modelul calității

Dezvoltarea QMF depinde de o definiție precisă a calității. Punctul de plecare este definirea calității rezultatului din perspectiva client/utilizator. Cea mai generală și succintă definiție a calității produsului este *adecvarea acestuia pentru utilizare*. Într-un mod mai detaliat, dar mai degrabă empiric, calitatea produsului este definită și de ISO 9000, care din perspectiva statisticii este sinonimă cu calitatea rezultatului, respectiv:

- Gradul în care un set de caracteristici intrinseci îndeplinește cerințele.

Definiția mai cuprinzătoare furnizată de ISO 8402:1986 este:

- Totalitatea trăsăturilor și caracteristicilor unui produs sau serviciu, care conferă abilitatea acestuia de a satisface nevoile exprimate și implicite.

Aceste definiții largi asigură o noțiune de bază a calității rezultatului, dar trebuie să fie însoțite de o interpretare mai precisă a *componentelor* calității. Acest lucru este prevăzut în *Definiția calității a SSE*, prezentată la întâlnirea din octombrie 2003 a Grupului de Lucru SSE *Evaluarea calității în statistică*. Deși nu i s-a conferit un statut mai formal, acest document a asigurat baza pentru definirea componentelor calității pentru toate documentele ulterioare ale SSE privind calitatea și a fost adoptat de către INS.

Componentele calității rezultatului

INS definește calitatea rezultatului în conformitate cu *Definiția calității a SSE*. În termeni mai largi, calitatea rezultatelor INS este evaluată în termenii "adecvării pentru utilizare". Mai precis, calitatea rezultatului se măsoară în termenii a șase componente ale calității: relevanța, acuratețe, oportunitatea și punctualitatea, claritatea și accesibilitatea, coerența și comparabilitatea.

Relevanța este gradul în care rezultatul statistic răspunde nevoilor curente și potențiale ale utilizatorului. Aceasta depinde de producerea tuturor statisticilor cerute de utilizatori și de gradul în care conceptele utilizate (definiții, clasificări, etc.) reflectă nevoile utilizatorilor. Utilizatorii au o varietate de nevoi, uneori contradictorii. Provocarea INS este de a cântări și de a echilibra nevoile actualilor sau potențialilor utilizatori prin proiectarea de anchete în limitele resurselor organizației.

Acuratețea rezultatului statistic în sens general este gradul de apropiere al estimațiilor de valorile reale. Aceasta se caracterizează, în general, în termenii estimării erorii statistice și în mod

tradițional poate fi împărțită în componente ale abaterii (eroare sistematică) și dispersie (eroare aleatoare). Poate fi de asemenea descrisă în termenii surselor majore de eroare, care pot determina lipsa acurateței (ex: acoperire, eșantionare, răspuns, non-răspuns și prelucrare).

Oportunitatea rezultatului statistic este durată de timp dintre momentul producerii evenimentului sau fenomenului descris și momentul în care acesta este disponibil. Informația care este disponibilă pentru utilizatori în timp util pentru scopul său principal, este considerată oportună. Deciziile privind oportunitatea implică compromisuri între relevanță și acuratețe.

Punctualitatea rezultatului statistic este decalajul dintre momentul furnizării datelor și data planificată pentru furnizare, după cum se anunță în calendarul oficial, stabilit prin reglementări sau acordat în prealabil între parteneri.

Accesibilitatea rezultatului statistic este ușurința cu care utilizatorii pot obține datele. Aceasta este determinată de condițiile fizice prin intermediul cărora utilizatorii obțin datele: unde să meargă, cum să comande, timpul livrării, politica de prețuri, condițiile de marketing (drepturile de autor, etc.), disponibilitatea micro sau macro datelor, diverse formate (hârtie, fișiere, CD-ROM, Internet, etc.)

Claritatea rezultatului statistic este măsura în care utilizatorii înțeleg cu ușurință datele. Aceasta este determinată de mediul informatic în care sunt prezentate datele, dacă datele sunt însoțite de metadatele adecvate, descriind conceptele, sursele, metodele, dacă utilizarea este prin ilustrații precum grafice și hărți, dacă informația privind acuratețea datelor este disponibilă (incluzând orice limitări în ce privește utilizarea) și măsura în care se acordă asistență suplimentară utilizatorilor.

Coerența a două sau mai multe rezultate statistice se referă la gradul în care procesele statistice, prin care au fost generate, au utilizat aceleași concepte - clasificări, definiții și populații țintă - și metode armonizate. Rezultatele statistice coerente au potențialul de a fi combinate și utilizate împreună. Exemple de o astfel de utilizare sunt atunci când rezultatele statistice se referă la aceeași populație, perioadă de referință și regiune, dar cuprind seturi diferite de date (de ex.: date de ocupare și date de producție) sau atunci când ele cuprind aceleași date (de ex.: datele de ocupare) dar pentru perioade de referință, regiuni sau alte domenii diferite.

Comparabilitatea este un caz special de coerență și se referă la ultimul exemplu, unde rezultatele statistice se referă la aceleași date, iar scopul combinării lor este de a putea face comparații în timp, pe regiuni sau pe alte domenii.

Componentele calității procesului

Calitatea rezultatului se obține prin *calitatea procesului*, care are două aspecte clare:

- *eficacitatea*: care conduce la rezultate de bună calitate;
- *eficiența*: care conduce la producerea acestora la costuri minime pentru INS, pentru respondenți și alți furnizori de date.

Componentele calității procesului sunt formulate în două grupe bazate pe primele zece principii din Codul de Practici al SSE, după cum urmează.

Mediul Instituțional

- *Independența profesională*: personalul INS este independent profesional de alte politici, alte departamente și organisme legislative sau administrative și de organizațiile din sectorul privat.
- *Mandat pentru culegerea de date*: INS are o împuternicire legală clară pentru culegerea informației. Furnizorii pot fi obligați prin lege să permită accesul la date sau să furnizeze date.
- *Adecvarea resurselor*: resursele disponibile sunt suficiente pentru a răspunde cerințelor sistemului și de prelucrare.

- *Angajamentul privind calitatea:* personalul se angajează să lucreze și să coopereze în conformitate cu principiile stabilite în Declarația de Calitatea a SSE.
- *Confidențialitatea statistică:* confidențialitatea furnizorilor de date (gospodării, întreprinderi, administrații și alți respondenți) și confidențialitatea informației pe care o furnizează și utilizarea acesteia doar pentru scopuri statistice sunt garantate.
- *Imparțialitatea și obiectivitatea:* producerea și diseminarea statisticii respectă independența științifică, iar statistica este realizată într-o manieră obiectivă, profesională și transparentă, prin care toți utilizatorii sunt tratați echitabil.

Anchetele și infrastructura statistică

- *Metodologie solidă:* instrumente, proceduri și expertiză adecvate.
- *Proceduri statistice adecvate:* sunt implementate de-a lungul ciclului de viață al anchetei.
- *Sarcină non-excesivă pentru respondenți:* sarcina de raportare este proporțională cu nevoile utilizatorilor și nu este excesivă. Este monitorizată de-a lungul timpului și sunt stabilite țintele privind reducerea acesteia.
- *Eficacitatea costurilor:* resursele sunt utilizate în mod eficient.

Considerații privind echilibrarea calității

Componentele calității rezultatului și procesului sunt interconectate. Nu există un model general care să le adune laolaltă pentru a optimiza sau pentru a prescrie un nivel al calității. Obținerea unui nivel acceptabil de calitate este rezultatul preocupării pentru realizarea unui echilibru între componentele calității de-a lungul timpului. Acest echilibru este un aspect foarte important al proiectării, implementării și funcționării anchetelor INS.

Capitolul 3 din acest ghid reprezintă un sprijin în promovarea bunelor practici statistice, în special a metodologiei clare, a procedurilor statistice adecvate, a sarcinii neexcesive asupra respondenților, a eficacității costurilor și a tuturor aspectelor calității rezultatului.

2.5 Evaluarea și raportarea calității

Metodele și procedurile SSE

- Standardul SSE privind rapoartele de calitate, care a fost actualizat în 2009, furnizează recomandări pentru pregătirea rapoartelor de calitate cuprinzătoare, de către INS-uri și direcțiile Eurostat.
- *Manualul SSE privind rapoartele de calitate*, de asemenea actualizat în 2009, asigură mai multe detalii și exemple.
- Ambele documente conțin cea mai recentă versiune a *Indicatorilor standard de calitate* pentru utilizare în centralizarea calității rezultatelor statistice.
- *Glosarul calității SSE* acoperă mai mulți termeni tehnici în documentația SSE privind calitatea, asigurând o scurtă definiție a fiecărui termen și indicând sursa definiției.
- Un glosar mai cuprinzător și mai actualizat este *Vocabularul comun de metadate*, dezvoltat în parteneriat cu organizațiile internaționale, inclusiv Eurostat.
- *Lista de control SSE pentru managerii de cercetări statistice (DESAP)* permite realizarea evaluărilor rapide, dar sistematice și cuprinzătoare ale calității unei anchete și a rezultatelor acesteia, precum și identificarea posibilelor îmbunătățiri.
- Există o versiune condensată a *Listei de control SSE pentru managerii de cercetări statistice (DESAP)*, care conține doar întrebări cheie selectate și există și o Versiune electronică a *Listei de control SSE* cu o Versiune electronică a *ghidului de utilizare*.

- *Manualul privind calitatea datelor – Metode și instrumente de evaluare (DatQAM)* detaliază întreaga gamă de metode pentru evaluarea calității procesului și rezultatului și instrumentele pe care se sprijină.
- *Manualul privind îmbunătățirea calității prin analiza variabilelor procesului* al SSE descrie o abordare generală și instrumente utile pentru identificarea, măsurarea și analizarea variabilelor cheie ale procesului.

Ultimele versiuni ale tuturor acestor documente sunt disponibile pe pagina de calitate a website-ului Eurostat. Luând în considerare abordările prezentate, mijloacele de raportare a calității de către INS sunt discutate în Capitolul 4.

3 Ghidul calității

Remarci introductive

Acest ghid are ca scop sprijinirea proiectării unei noi anchete, reproiectării sau îmbunătățirii celor deja existente. Acesta este organizat în secțiuni corespunzătoare principalelor sub-procese din care este alcătuită în mod tipic o anchetă. Aceasta pare a fi o structură mai naturală pentru managerii de anchetă (utilizatorii principali) decât organizarea aceleiași informații în funcție de componentele calității rezultatului și proceselor. S-a acordat de asemenea atenție utilizării aceleiași structuri precum recentul promulgat model generic al procesului statistic. Totuși, acest model își are originea într-un mediu IT și include o distincție exactă între funcțiile de proiectare și construcție, în timp ce, din perspectiva unui manager de anchetă este mai adecvat să spunem tot ce e de spus despre un anumit sub-proces, de exemplu eșantionarea, în cadrul unei singure secțiuni, indiferent dacă se leagă de proiectare, construcție sau operare.

Secțiunea 3.1 furnizează recomandări de asistență în identificarea obiectivelor, utilizărilor și utilizatorilor. Secțiunile ce urmează acoperă sub-procesele ulterioare ale anchetei în ordine aproximativ cronologică. Totuși, unele procese pot avea loc simultan, de exemplu, dezvoltarea cadrului, eșantionarea și proiectarea chestionarului. Există interdependențe semnificative între unele sub-procese, de exemplu, între proiectarea chestionarului, colectarea și procesarea datelor. Alte secțiuni, precum documentarea, se referă la activități care nu constituie sub-procese distincte de sine stătătoare, ci au mai degrabă loc în contextul mai multor sub-procese. În cele din urmă, aspectele legate de utilizarea datelor administrative sunt discutate într-o secțiune separată.

Toate secțiunile sunt structurate în patru părți, după cum urmează:

- *Scopul și obiectivele*, cuprinzând o definiție și descrierea sub-procesului sau activității și o indicație a impactului său potențial asupra calității, stabilind astfel contextul pentru următoarele recomandări.
- *Principiile*, cuprinzând politicile și abordările largi care guvernează proiectarea sub-procesului sau a activității, cu accent pe acelea care se referă la calitate.
- *Recomandările*, cuprinzând bunele practici cunoscute și furnizând o listă tehnică de control cuprinzătoare în sprijinul proiectării și derulării anchetei. Este nevoie de discernământ pentru a decide dacă practicile sunt potrivite pentru fiecare caz în parte și momentul și modalitatea în care acestea trebuie implementate.
- *Referințe*. Sunt prezentate standarde și politici internaționale sau ale INS, relevante, împreună cu documentele care furnizează detalii privind anumite aspecte.

3.1 Obiective, utilizări și utilizatori

Scop și obiective

Înainte de aprobare, trebuie efectuat un exercițiu inițial de planificare în vederea stabilirii nevoilor principale de date și a fezabilității generale a unei anchete nou propuse sau reproiectate. Acesta include definirea obiectivelor generale, populația țintă și întrebările sau aspectele cheie, către care se va îndrepta analiza bazată pe date. După aprobare, trebuie elaborate mai departe rezultatele planificării inițiale pentru a asigura că activitatea nouă sau reproiectată va respecta cerințele specifice ale utilizatorului.

Obiectivele sunt scopurile pentru care este cerută informația, enunțate în contextul programului, problema de cercetare sau ipotezele care dau naștere nevoii de date.

Utilizările specifice indică mai precis nevoia de date, de pildă, descriind ce decizii pot fi luate pe baza datelor culese și în ce măsură sprijină informațiile aceste decizii. Pentru anchetele periodice, de-a lungul timpului se pot dezvolta alte utilizări.

Utilizatorii sunt organizațiile și indivizii care urmează să utilizeze informația.

Ajungerea la un consens asupra obiectivelor specifice, utilizărilor și utilizatorilor, este esențială pentru luarea de decizii raționale în proiectarea anchetei.

Principii

Specificarea obiectivelor și a utilizărilor prin consultarea cu potențialii utilizatori, conduce la dezvoltarea unui plan detaliat al anchetei. Utilizatorii pot ajuta la descrierea mai precisă a obiectivelor anchetei. Dezvoltatorii de anchete pot identifica aspectele conceptuale, metodologice și operaționale pe care trebuie să le rezolve și pot sugera un program rezonabil.

Este important să avem o înțelegere clară și să formulăm concret obiectivele în termenii ipotezelor ce vor fi testate și/sau ai cerințelor specifice privind datele, inclusiv calitatea așteptată, constrângerile bugetului și datele de livrare așteptate.

O formulare a obiectivelor este un mijloc important de comunicare către utilizatori a rezultatelor așteptate și a limitărilor acestora. Aceasta asigură, de asemenea, utilizatorilor viitori, care pot avea obiective diferite, mijloacele necesare evaluării gradului în care rezultatele pot răspunde nevoilor lor specifice.

Recomandări

- Focalizarea analizei asupra nevoilor utilizatorului de a găsi cele mai eficiente soluții privind costurile, atât pe termen lung, cât și pe termen scurt. Înainte de a începe proiectarea/reproiectarea, analizați statistica actuală disponibilă din domeniu în ceea ce privește sursele, frecvența, calitatea, oportunitatea, etc. Aveți în vedere compromisul dintre limitările datelor disponibile și costurile și timpul necesare pentru producerea datelor suplimentare.
- Dezvoltați obiectivele anchetei în parteneriat cu utilizatorii și părțile interesate importante. Identificați utilizatorii și grupurile de utilizatori importanți, de exemplu pe cei care elaborează politicile. Includeți potențialii utilizatori din interiorul INS, de exemplu Conturile Naționale. Stabiliți și mențineți relațiile cu utilizatorii de informație din sectorul public și privat și cu publicul general, pentru a întări relevanța rezultatelor și pentru a îmbunătăți diseminarea lor.
- Înainte de orice proiectare sau reproiectare majoră, stabiliți consultări cu utilizatorii, pentru a putea identifica opțiunile, precum și pentru a putea obține sprijinul publicului pentru program. Întrucât rezultatele nu sunt folosite dacă nu inspiră încredere, apălați la o abordare deschisă atunci când dezvoltați sau revizuiți programe, astfel încât să puteți construi încrederea în procese și produse.
- În determinarea măsurii în care o anchetă va răspunde nevoilor utilizatorului, căutați un compromis rezonabil între aceste nevoi și buget, între sarcina de răspuns și aspectele privind confidențialitatea. INS poate avea puțin de spus atunci când metodologia generală și conținutul sunt legiferate. Cu toate acestea, în alte cazuri, merită să formulăm abordări metodologice alternative, mijloace și moduri de culegere a datelor, frecvențe, detalii geografice, etc., în vederea atingerii soluției optime.
- Revizuirea anchetelor continue la intervale regulate. Anchetele trebuie să evolueze, să se adapteze și să fie reînnoite, astfel încât să poată menține pasul cu cererile în schimbare ale utilizatorilor. Scopul anchetelor trebuie revizuit și reexaminat periodic pentru a spori relevanța vizavi de nevoile utilizatorului, întrucât acestea pot evolua sau se pot modifica. Uneori poate fi necesară modificarea anchetelor existente pentru menținerea fiabilității seriilor statistice cheie, în special dacă sursele de informație s-au schimbat sau în cazul în care mijloacele de obținere au fost reproiectate.
- Atunci când există ținte explicite privind calitatea datelor, includeți-le în formularea obiectivelor anchetei. Țintele se pot referi la erorile de eșantionare, ratele de acoperire, ratele de răspuns și oportunitatea.

Referințe

Blanc, M., Radermacher, W. and Körner, T. (2001). *Quality and users*. Session 15.1 of the International Conference on Quality in Official Statistics, Stockholm, Sweden.

Brackstone, G.J. (1993). *Data relevance: keeping pace with user needs*. Journal of Official Statistics, 9, 49-56.

3.2 Concepte, variabile și clasificări**Scop și obiective**

Conceptele sunt subiectele interpelărilor și analizelor de interes ale utilizatorilor. Acestea se referă la caracteristicile generale sau atributele unităților statistice. *Variabilele* sunt indicatori folosiți pentru a măsura conceptele. Definițiile conceptelor și variabilelor corespunzătoare trebuie să fie clare, specificate precis și relevante pentru scopurile analitice primare, pentru care sunt culese datele. Toate categoriile de date statistice trebuie să fie clasificate în vederea prezentării și analizei, iar *schema de clasificări* corespunzătoare trebuie să fie potrivită pentru aceste scopuri analitice.

Principii

Pentru a trage concluzii dintr-un set de date, utilizatorii trebuie să aibă acces și să cunoască conceptele pe care se fundamentează datele. Utilizarea definițiilor armonizate ale conceptelor, variabilelor și clasificărilor ajută utilizatorii în compararea și integrarea datelor. Totuși, astfel de definiții pot avea nevoie de modificări pentru a răspunde nevoilor specifice ale utilizatorilor. O variabilă aleasă la un moment dat poate deveni perimată dacă sunt angrenați noi factori.

Definițiile conceptelor, variabilelor și clasificărilor trebuie documentate și orice diferențe față de standarde sau față de definițiile utilizate pentru datele asociate trebuie scoase în evidență. Această documentație este în mod special importantă pentru utilizatorii care doresc să folosească datele în alte scopuri decât cele prevăzute inițial.

Există adesea mai multe moduri de a măsura un concept. Factorii ce trebuie luați în considerare în definirea indicatorilor includ ușurința în obținerea datelor de la respondenți, sarcina de răspuns impusă respondentului, contextul în care trebuie puse întrebările, metoda de culegere, editarea, imputarea și tehnicile de ponderare, dacă informația poate fi obținută din rapoartele administrative și costurile asociate cu culegerea și prelucrarea.

Recomandări

- Specificați conceptele, variabilele și clasificările în mod clar și justificați-le în funcție de scopul prevăzut.
- Unde este posibil, utilizați definițiile standard ale conceptelor, variabilelor, clasificărilor, unităților statistice și populațiilor.
- În alegerea denumirilor convenționale, luați în considerare standardele și utilizările existente. Folosiți titlurile din standardele existente doar așa cum sunt definite în standarde.
- Pentru a maximiza flexibilitatea utilizării, codificați microdatele și mențineți fișierele la cel mai jos nivel posibil al fiecărei clasificări. Agregarea la un nivel mai înalt, cerută pentru anumite scopuri analitice sau pentru a satisface constrângerile legate de confidențialitate sau fiabilitate a datelor, se poate realiza într-o etapă ulterioară.
- Unde este posibil, folosiți o strategie comună pentru agregări și definiți-le în termenii agregărilor de la un nivel superior al clasificărilor. Documentați diferențele dintre nivelurile standard și cele afective ale clasificărilor/agregărilor utilizate. Folosiți clasificările care includ nivelurile detaliate și pe cele „în cascadă”. Specificați clar utilizatorilor cum se încadrează acestea în clasificările de la un nivel superior (adică, mai puțin detaliate).

- Folosiți definițiile standard pentru a face posibilă compararea datelor culese din surse diferite și integrarea datelor pe surse. INS are multe clasificări standard, inclusiv privind ramurile industriale, produsele, ocupațiile și geografia (vezi referințele de mai jos), precum și un glosar al termenilor statistici. În plus, există clasificări standard internaționale produse de Oficiul Statistic al Națiunilor Unite, Oficiul Internațional al Muncii, Eurostat și alte agenții internaționale și regionale. În absența unui standard oficial, examinați conceptele, variabilele și clasificările care sunt folosite de programele statistice aferente și consultați-vă cu experții în standarde, acolo unde este necesar.
- INS a realizat concordanța cu mai multe clasificări internaționale standard. Când există cerințe de date din partea agențiilor internaționale, folosiți concordanțele oficiale, când acestea sunt disponibile.
- Utilizați unitățile standard de observare pentru a facilita compararea datelor. Clasificările sunt de obicei proiectate cu luarea în considerare a anumitor unități de observare.
- Luați în considerare activitățile statistice derivate, precum Sistemul Conturilor Naționale, ale căror cerințe în termenii variabilelor și clasificărilor pot avea un efect semnificativ asupra anchetei.

Referințe

Clasificarea activităților din Economia Națională - CAEN REV.1
 Clasificarea activităților din Economia Națională - CAEN REV.2
 Clasificarea Produselor și Serviciilor asociate Activităților din Economia Națională – CPSA
 Nomenclatorul Produselor și Serviciilor Industriale – PRODRUM
 Registrul unităților administrativ-teritoriale a României – SIRUTA
 Nomenclatorul Unităților Teritoriale Statistice – NUTS
 Clasificarea Ocupațiilor din România – COR 2009
 Clasificarea Internațională Standard a Educației – ISCED
 Inventarele naționale de emisii – CORINAIR
 Nomenclatorul Selectat pentru Sursele de Poluare a Aerului – SNAP
 Clasificarea standard pentru statistica transporturilor - NST 2007
 Nomenclatorul mărfurilor periculoase
 Nomenclatorul de Produse și Servicii – COICOP
 Nomenclatorul Combinat de Bunuri – NC
 Sistemul Conturilor Naționale, 1995

3.3 Acoperire și cadrul de eșantionare

Scop și obiective

Populația țintă este setul de unități pentru care se cer informații și sunt cerute estimări. Considerentele practice pot dicta ca unele unități să fie în mod deliberat excluse, de exemplu, micile întreprinderi, care sunt imposibil de accesat fără a implica costuri mari. De aici se naște conceptul de *populație statistică*, setul de unități obținut prin excluderea unor unități care sunt greu sau costisitor de accesat. Diferențele dintre populația țintă și populația statistică sunt rezultatul restricțiilor deliberate ale acoperirii. Dacă cele două diferă, se poate face o inferență statistică validă sub eșantionarea probabilistică asupra populației statistice, și nu asupra populației țintă.

Cadrul este orice listă, material sau dispozitiv care delimitează, identifică și permite accesul la elementele populației anchetate. Cadrele sunt în general de două tipuri: cadre zonale și liste de cadre. Cadrele zonale sunt de obicei alcătuite dintr-o ierarhie de unități geografice, respectiv, unitățile de la un anumit nivel al cadrului pot fi subdivizate pentru a forma unitățile de la următorul nivel. Toate elementele incluse în cadrul constituie *populația-cadru*. Discrepanțele dintre populația cercetată și populația-cadru sunt denumite *erori de acoperire*.

Principii

Cadrul anchetei trebuie să se conformeze populației cercetate și să conțină subacoperirea și supraacoperirea minimală. Informația cu privire la cadre trebuie menținută la zi. Erorile de acoperire apar din cauza omisiunilor, includerilor eronate, dublărilor și/sau clasificărilor eronate ale unităților din cadrul eșantionului.

Caracteristicile unităților din cadru (de ex.: identificare, contact, clasificare, adresă, dimensiune, hărți, în cazul unităților geografice), trebuie să fie de înaltă calitate în vederea utilizării lor în stratificare, selecția de eșantion, culegere, urmărire, prelucrare de date, imputare, estimare, legăturile între înregistrări, evaluarea calității și analiză. Imperfecțiunile cadrului precum erorile de acoperire și caracteristicile depășite pot abate sau diminua fiabilitatea estimărilor anchetei sau pot crește costurile generate de culegerea de date.

Recomandări

- Testați cadrele posibile în stadiul de planificare a anchetei pentru identificarea gradului lor de adecvare și calității lor.
- Când există mai multe cadre, unele dintre acestea fiind incomplete dar mai puțin costisitoare pentru utilizare și altele mai complete dar în mod prohibitiv costisitoare, luați în considerare utilizarea cadrelor multiple.
- Luați în considerare că se poate utiliza apelarea aleatoare a numerelor în anumite anchete telefonice în combinație cu alte cadre zonale sau liste.
- Pentru anchetele în instituții și în întreprinderi, utilizați Registrul Statistic al Întreprinderilor. (REGIS)
- Pentru anchetele în agricultură, utilizați Registrul Statistic al Exploatațiilor Agricole (REXA).
- Pentru anchetele în gospodării, folosiți Eșantionul Master pentru Anchetele în Gospodării (EMZOT).
- Pentru a evita contradicțiile, pentru a facilita combinarea estimărilor din anchete și pentru a reduce costurile menținerii cadrului și evaluării, utilizați același cadru pentru anchete cu aceeași populație țintă, pe cât posibil.
- Rețineți și stocați informația despre eșantionare, rotația și culegerea de date astfel încât să se poată obține coordonarea dintre anchete, iar relațiile respondenților și sarcina de răspuns să poată fi mai bine gestionate. De exemplu, înregistrați pe unități datele culese pentru fiecare anchetă.
- Încorporați procedurile pentru a elimina dublarea și pentru a actualiza informațiile privind nașterile, decesele unităților din afara domeniului și modificările în caracteristici.
- Monitorizați calitatea cadrului evaluând periodic acoperirea sa și calitatea informației privind caracteristicile unităților. Există mai multe tehnici în acest scop:
 - Potrivirea cadrului sau a eșantionului unui cadru cu sursele alternative comparabile, furnizate adesea de înregistrările administrative, pentru populația anchetată sau subseturile acesteia;
 - Analizarea rezultatelor anchetelor în vederea depistării duplicărilor, deceselor, unităților din afara domeniului și modificărilor caracteristicilor;
 - Utilizarea întrebărilor specifice în chestionare pentru a sprijini monitorizarea informațiilor privind acoperirea și clasificarea; verificarea informațiilor împreună cu

autoritățile locale (ex.: oficiile regionale, personalul care asigura ancheta pe teren, unitățile de anchetă);

- Verificarea cadrului sau a subseturilor acestuia pe teren (aceasta poate include verificarea unităților din afara domeniului);
- Compararea cadrului cu eșantionul unităților dintr-un cadru zonal corespondent;
- Actualizarea cadrului pentru determinarea modificărilor acestuia;
- Verificarea consistenței înregistrărilor cu alte surse sau cu datele din surse special concepute;
- Utilizarea informației obținute din alte anchete cu același cadru pentru evaluare;
- Monitorizați cadrul dintre momentul selectării eșantionului și perioada de referință a anchetei.
- Încorporați actualizările cadrului în maniera cea mai oportună cu putință.
- Minimizați erorile cadrului prin instruirea eficace a personalului, cu accent pe importanța acoperirii și pe implementarea procedurilor de asigurare a calității activităților legate de cadru.
- Pentru cadrele zonale, implementați verificarea hărților pentru a asigura delimitarea clară și fără suprapuneri a zonelor geografice utilizate în proiectarea eșantionului (de ex.: prin verificări pe teren sau prin utilizarea altor surse de hărți).
- Pentru activitățile statistice din sursele administrative sau pentru activitățile statistice derivate, unde modificările acoperirii pot ieși de sub controlul managerului responsabil, determinați și monitorizați acoperirea și negociați modificările necesare cu managerul sursei.
- Faceți ajustări ale datelor sau utilizați date suplimentare din alte surse pentru a compensa eroarea de acoperire a cadrului.
- Includeți în documentația anchetei descrieri ale populațiilor țintă și ale celor anchetate, orice diferențe dintre populația țintă și populația anchetată, precum și descrierea cadrului și a erorilor sale de acoperire.

Referințe

Archer, D. (1995). *Maintenance of business registers*. In Business Survey Methods, B.G. Cox et al (eds.), Wiley-Interscience, New York, 85-100.

Colledge, M.J. (1995). *Frames and business registers: an overview*. In Business Survey Methods, B.G. Cox et al (eds.), Wiley-Interscience, New York, 21-47.

Eurostat (2003) *Business Register Recommendations Manual* (http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-BG-03-001/EN/KS-BG-03-001-EN.PDF)

Kott, P.S. and Vogel, F.A. (1995). *Multiple-frame business surveys*. In Business Survey Methods, B.G. Cox et al (eds.), Wiley-Interscience, New York, 185-203.

3.4 Eșantionarea

Scop și obiective

Eșantionarea este selectarea unui set de unități dintr-o populație anchetată, set denumit eșantion, având ca scop elaborarea de prezumpții despre populație. Alegerea metodei de eșantionare are o influență considerabilă asupra calității, în special asupra erorilor de eșantionare. Depinde de mai mulți factori, incluzând nivelul dorit de precizie, gradul de detaliu al datelor de ieșire, disponibilitatea cadrului potrivit de eșantionare și a variabilelor auxiliare

pentru stratificare și selecția eșantionului, metodele de estimare ce vor fi folosite și bugetele disponibile.

Principii

Eșantionarea aleatorie reprezintă selectarea unităților dintr-un cadru astfel încât toate unitățile să aibă *probabilități de includere* pozitive și cunoscute. Dimensiunea eșantionului este determinată de precizia cerută și bugetul disponibil. Distribuția probabilității care guvernează selecția eșantionului, împreună cu etapele și unitățile de eșantionare, stratificarea, determinarea dimensiunii eșantionului și alocarea sunt denumite colectiv *planul de eșantionare*. Se alege o combinație de plan de eșantionare și metodă de estimare (vezi Secțiunea 3.10), astfel încât estimările rezultate să aibă cea mai bună precizie în condițiile unui buget dat sau astfel încât să rezulte cele mai joase costuri pentru gradul de precizie stabilit.

Stratificarea constă în împărțirea populației în subseturi (*straturi*) în interiorul fiecăreia dintre ele fiind selectat un eșantion independent. Alegerea straturilor este determinată de obiectivele anchetei, de distribuțiile frecvenței variabilelor principale de interes și de precizia dorită.

Uneori, informația necesară pentru stratificarea populației nu este disponibilă în cadru. În astfel de cazuri, se poate folosi o schemă *de eșantionare în două etape (dublă)*, pentru obținerea datelor necesare pentru stratificare, acolo unde este selectat un eșantion mai mare într-o primă etapă. Primul eșantion este apoi stratificat conform cerințelor, iar eșantionul din cea de-a doua etapă este selectat în eșantionul din prima etapă.

În cazul cadrelor zonale, nu este eficientă selectarea directă a unităților *de raportare*, adică, unitățile pentru care vor fi raportate date. În astfel de cazuri, se poate utiliza o schemă *de eșantionare în două etape*. În prima etapă sunt selectate *clusterelor* (denumite *unități primare de eșantionare*) unităților potențiale de raportare; în cea de-a doua etapă, sunt selectate eșantioane ale unităților de raportare din fiecare unitate primară de eșantionare selectată. Pot fi necesare mai multe etape datorate constrângerilor bugetare sau de alt tip.

Datele culese pentru unitățile de raportare pot fi completate, acolo unde este cazul, în stadiul de estimare, cu *informații auxiliare* din alte surse decât ancheta în sine, precum înregistrările administrative sau proiecțiile de recensământ, pentru a îmbunătăți precizia estimărilor. În alegerea planului de eșantionare trebuie să se țină cont de disponibilitatea acestor informații auxiliare.

Recomandări

- Pentru eficiența statistică, straturile ar trebui să conțină unități cât mai omogene cu putință, din perspectiva informației cerute în anchetă.
- Pentru populațiile cu un grad mare de asimetrie, care caracterizează majoritatea populațiilor de întreprinderi, creați un strat de unități mari care să fie incluse în anchetă. Aceste unități mari au un impact semnificativ asupra estimărilor totalurilor populației.
- Majoritatea anchetelor produc estimări pentru diverse *domenii* de interes, de exemplu, regiunile. Acest lucru trebuie luat în considerare în proiectare prin stratificarea în conformitate cu aceste domenii, unde este posibil; în caz contrar, este necesar să se efectueze post-stratificarea în etapa estimării.
- În cazul anchetelor longitudinale, alegeți variabilele de stratificare care corespund caracteristicilor stabile de-a lungul timpului.
- Când nu sunt disponibile din cadru toate datele cerute pentru o bună stratificare, luați în considerare posibilitatea unui plan de eșantionare în două etape. Calculați costul de eșantionare la fiecare etapă, disponibilitatea datelor cerute la fiecare etapă și ce se câștigă în privința preciziei obținute prin stratificarea eșantionului din prima etapă.
- Pentru anchetele bazate pe cadrul zonal, determinați câte etape de eșantionare sunt necesare și care dintre unitățile de eșantionare sunt potrivite la fiecare etapă. Pentru fiecare tip posibil de unitate de raportare, aveți în vedere disponibilitatea obținerii sau creării unui

cadru potrivit al acestor unități, ușurința contactării și culegerii de date, calitatea datelor și costul culegerii datelor.

- Elaborați studii pentru evaluarea metodelor alternative de eșantionare, opțiunile de stratificare și posibilitățile de alocare. Utilitatea acestor studii depinde de disponibilitatea și vechimea datelor utilizate pentru studii și relația lor cu indicatorii principali ai anchetei. Aceste date pot fi extrase din recensămintele anterioare, din anchete sau date administrative.
- La determinarea dimensiunii eșantionului, aveți în vedere nivelurile de precizie cerute pentru fiecare indicator, planul de eșantionare, informațiile auxiliare, bugetul și factorii de non-eșantionare precum non-răspunsul, unitățile din afara domeniului și uzura (din anchetele longitudinale și cele periodice).
- Majoritatea anchetelor produc estimări pentru mulți indicatori, iar eșantionul care este optimizat pentru unul dintre aceștia, poate fi departe de a fi optimizat și pentru ceilalți. Gestionați această problemă identificând mai întâi indicatorii cheie și apoi optimizați în raport cu unul sau doi din acest subset. Acest lucru necesită un compromis între strategiile optime pentru indicatorii individuali.
- În determinarea alocării eșantionului și a dimensiunii eșantioanelor stratificate, luați în considerare deficiențele de cadru așteptate, precum unitățile decedate și greșit clasificate.
- În etapa implementării, comparați dimensiunea și caracteristicile eșantionului actual cu ceea ce s-a așteptat. Comparați precizia estimărilor cu obiectivele planificate.

Anchetele periodice

- Întrucât populația anchetată dintr-o anchetă periodică poate crește treptat, este prudent să folosim un plan de eșantionare, care menține dimensiunea eșantionului (și de aici costurile de culegere) stabilă. Acest lucru este posibil întrucât precizia estimărilor anchetei este mai dependentă de dimensiunea totală a eșantionului, decât de proporția pe care o reprezintă eșantionul din total.
- Proiectarea trebuie să fie cât mai flexibilă cu putință pentru a preîntâmpina viitoarele schimbări, precum creșterile sau scăderile dimensiunii eșantionului, restratificarea, reeșantionarea și actualizarea probabilităților de selecție. Dacă se cer estimări pentru domenii de interes specifice (ex.: estimări subprovinciale), formați straturile combinând domenii stabile mai mici (ex.: sectoare), astfel încât schimbările viitoare în domeniile cerute să fie mai ușor de adaptat.
- Dacă se cer estimări eficiente de schimbare și/sau dacă sarcina de răspuns este o preocupare, folosiți schema de eșantionare *rotațională*, care înlocuiește o parte a eșantionului în fiecare perioadă. Alegerea ratei de rotație va fi un compromis între precizia estimărilor și sarcina respondentului. Scăderea ratei de rotație asigură o bună precizie, dar în timp pot scădea ratele de răspuns. Aceasta beneficiază suplimentar de reducerea costurilor dacă primul contact este în mod substanțial mai scump decât contactele ulterioare.
- Dezvoltați procedurile de monitorizare a calității proiectării eșantionului în timp. Stabiliți o strategie actualizată pentru re-proiectarea selectivă a straturilor care au suferit o deteriorare serioasă.
- Pentru anchetele panel longitudinale, determinați lungimea perioadei panelurilor în eșantion, echilibrând durata cu efectele de uzură și cele de condiționare. Folosiți un model cu paneluri suprapuse (adică, cu o durată de timp suprapusă) acolo unde este nevoie să se producă estimări transversale împreună cu cele longitudinale.

Software

- Folosiți software generic în locul sistemelor personalizate ori de câte ori este posibil pentru a reduce costurile de dezvoltare și posibilele erori de programare. De exemplu MICROSTRATE, dezvoltat de Eurostat, permite controlul suprapunerilor eșantioanelor.

Referințe

Cochran, W.G. (1977). *Sampling Techniques*. Wiley, New York.

Eurostat (2008) *Eurostat Sampling Reference Guidelines – Introduction to Sample Design and Estimation Techniques*. (http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-003/EN/KS-RA-08-003-EN.PDF)

Eurostat (2006) *Benchmarking through Calibration of Weights for Microdata* (http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/KS-DT-05-007/EN/KS-DT-05-007-EN.PDF)

Hidiroglou, M.A., (1994). *Sampling and estimation for establishment surveys: stumbling blocks and progress*. Proceedings of the Section on Survey Research Methods, American Statistical Association, 153- 162.

Hidiroglou, M.A, and Srinath, K.P. (1993). *Problems associated with designing sub-annual business surveys*. Journal of Economic Statistics, 11, 397-405.

Kish, L. (1965). *Survey Sampling*. Wiley, New York.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

3.5 Proiectarea chestionarelor

Scop și obiective

Chestionarul este un set de întrebări proiectate să colecteze informația de la respondenți. Acesta poate fi administrat de operatorul de interviu sau completat de către respondent, folosindu-se metoda de culegere a datelor pe hârtie, metodele telefonice sau electronice, sau prin Internet. Chestionarele au un impact major asupra comportamentului respondentului, performanțelor operatorului de interviu, costurilor de culegere a datelor și relațiilor cu respondentul.

Principii

Proiectarea chestionarelor (mod, conținut și format) trebuie să ia în considerare cerințele utilizatorilor, operațiunile anchetei, precum și natura și caracteristicile populației respondente. Cele mai bune chestionare impun o sarcină redusă de răspuns și vin atât în sprijinul respondentului, cât și în cel al operatorului de interviu. Acestea conțin întrebări relevante, permit colectarea eficientă a datelor, cu un număr minim de erori, facilitând codificarea și captarea de date și minimizând ulterior editarea și imputarea.

Toate chestionarele noi și cele modificate trebuie testate amănunțit înainte de implementare. Chestionarele din anchetele permanente trebuie evaluate periodic.

Recomandări

- Utilizați în chestionare cuvinte și concepte care sunt ușor de înțeles de către respondenți. În cazul întreprinderilor, alegeți întrebările, perioadele de referință și categoriile de răspuns care sunt compatibile cu practicile contabile ale companiei respective. Armonizați conceptele și formulările, cu acelea aflate deja în uz. Refolosiți, acolo unde este cazul, întrebări din alte anchete.
- Alegeți modelul de întrebare și formularea care încurajează respondentul să completeze chestionarul cât mai precis cu putință. În acest scop, chestionarul nu trebuie să devieze de la subiectul anchetei, trebuie să fie cât mai scurt cu putință, să treacă cu ușurință de la o întrebare la alta, să faciliteze memoria respondentului și să indice spre cea mai adecvată sursă de informație.

- În introducerea chestionarului indicați subiectul anchetei, identificați sponsorul, explicați scopul anchetei și solicitați cooperarea respondentului. Indicați de asemenea autoritatea prin care se efectuează ancheta, ce măsuri de protejare a confidențialității se iau și orice acord stabilit în privința legăturilor între date sau partajarea acestora. Întrebările din introducerea chestionarului, trebuie să fie aplicate tuturor respondenților, trebuie să stabilească dacă respondentul este membru al populației anchetei și să fie ușor de completat.
- Proiectați chestionare completate prin auto-înregistrare de către respondenți care să fie atractive și ușor de completat. În acest scop, creați o primă impresie pozitivă în scrisoarea de intenție și în prima pagină și asigurați o notă de profesionalism a chestionarelor.
- Dacă chestionarul va fi administrat de către operatorul de interviu, creați-l spre a fi ușor de folosit de către acesta.
- Pentru a minimiza posibilitatea erorilor de raportare, asigurați-vă că instrucțiunile către respondenți și/sau operatorii de interviu sunt scurte, clare și ușor de găsit. Furnizați definiții la începutul chestionarului sau în întrebările specifice, după cum se cere (nu într-o foaie separată de instrucțiuni). Asigurați-vă că perioadele de referință și unitățile raportoare sunt clar definite. Utilizați îngroșarea caracterelor pentru a scoate în evidență elementele importante. Specificați "include" sau "exclude" în întrebări și asigurați-vă că acele categorii de răspuns se exclud reciproc și sunt cuprinzătoare.
- În legătură cu forma chestionarului, furnizați titluri pentru fiecare secțiune și instrucțiuni și spațiu pentru răspunsuri care să faciliteze acuratețea. Utilizați culori, evidențieri, ilustrații și simboluri pentru a atrage atenția și pentru a îndruma respondenții (sau pe operatorii de interviu) în privința chestionarului. La finalul chestionarului, puneți la dispoziție spațiu pentru comentariile suplimentare ale respondenților, luați în considerare posibilitatea de a întreba cât timp a durat completarea chestionarului și exprimați-vă aprecierea la adresa respondentului.

Testarea

- Cele mai adecvate metode de testare și evaluare a chestionarelor depind de tipul și dimensiunea anchetei, conținutul anchetei, măsura în care se utilizează întrebările anchetei anterioare, dacă ancheta este periodică sau nu, de metoda culegerii de date, programul de dezvoltare, buget și disponibilitatea resurselor.
- Aveți în vedere două sau mai multe etape de testare a chestionarelor, de exemplu, testarea chestionarelor la un stadiu primar de dezvoltare și apoi retestarea lor, dacă s-au efectuat revizuirii pe baza constatărilor de la testul inițial. Acest proces se poate repeta de mai multe ori, cu diferite metode de fiecare dată.
- Utilizați grupurile de discuții și interviuri în profunzime, interviuri cu opinii exprimate și codificarea comportamentului pentru a putea observa reacția respondenților la chestionar. Grupurile de discuții și cele unu-la-unu, interviurile în profunzime sunt folosite pentru testarea și evaluarea formulării, secvențierii și formatului întrebărilor. Metodele cognitive sunt folosite pentru a examina procesele de gândire a respondenților pe măsură ce răspund la întrebările din anchetă și pentru a examina dacă aceștia au înțeles sau nu întrebările și dacă pot astfel să dea răspunsuri precise. Codificarea comportamentului asigură mijloace sistematice și obiective de examinare a eficacității chestionarului, analizând interacțiunea operator interviu-respondent. Aceste metode de testare calitativă pot ajuta în determinarea conținutului chestionarului prin evaluarea și explorarea conceptelor cheie.
- Efectuați teste informale (pretestare) asupra chestionarelor pentru a sprijini identificarea ordonării sau formulării defectuoase a întrebărilor, erorilor în expunerea chestionarelor sau a instrucțiunilor și problemelor ce ar putea fi generate de imposibilitatea respondentului sau caracterul refractar al acestuia de a răspunde la întrebări. Folosiți testarea informală pentru a sugera categoriile suplimentare de răspuns care pot fi pre-codificate și pentru a indica durata preliminară a lungimii interviului.

- Organizați sesiuni de informare cu operatorii după testarea chestionarelor, în care aceștia discută asupra experiențelor lor cu respondenții intervievați și își exprimă părerea asupra percepției chestionarului. Operatorii pot identifica sursele potențiale de erori de răspuns și non-răspuns, precum și ariile unde chestionarul poate fi îmbunătățit.
- Folosiți testarea pe eșantioane divizate, atunci când este necesară determinarea celei mai bune versiuni ale chestionarului. Aceasta implică o proiectare experimentală care este încorporată în procesul de culegere a datelor. Aceasta poate fi folosită pentru investigarea problemelor în formularea întrebărilor, ordonarea întrebărilor și procedurile de culegere a datelor.
- Efectuați testarea pilot pentru a observa modul în care operațiunile anchetei, inclusiv administrarea chestionarului, conlucrează în practică. Testul pilot se efectuează cu versiunea finală a cercetării proiectată la scară mică de la început până la sfârșit, inclusiv prelucrarea și analiza de date. Acesta oferă oportunitatea de a ajusta chestionarul înainte de utilizarea lui.

Referințe

Fowler, F.J. Jr. (1995). *Improving Survey Questions: Design and Evaluation*. Applied Social Research Methods Series, 38, Sage Publications, Thousand Oaks, California.

Gower, A.R. (1994). *Questionnaire design for business surveys*, Survey Methodology, 20, 125-136.

3.6 Culegerea și captarea de date

Scop și obiective

Culegerea de date reprezintă procesul de obținere a datelor de la indivizii sau organizațiile care folosesc chestionarul. Dacă la început nu se obține nicio dată sau dacă datele obținute sunt incomplete sau nepotrivite (conform editării preliminare), atunci se poate cere recontactarea pentru a sprijini finalizarea culegerii datelor (după cum se va discuta în următoarea secțiune).

Captarea de date este procesul care transformă informațiile furnizate de respondenți într-un format electronic adecvat prelucrării ulterioare. În cazul CAPI, CATI și EDR, datele sunt capturate ca parte a procesului efectiv de culegere. În alte cazuri, datele sunt capturate printr-o operațiune separată prin introducere manuală sau prin mijloace automate precum Recunoașterea Inteligentă a Caracterelor (ICR). Adesea, această conversie a datelor implică fie codificarea manuală, fie automată. Uneori, aceasta include transmiterea datelor către o altă locație.

Principii

Furnizorii de date, în special indivizii și organizațiile care completează chestionarele, de obicei fără a fi remunerați, reprezintă resursa cea mai valoroasă a INS. Pentru asigurarea unei cooperări continue, este esențial ca sarcina de răspuns să fie minimalizată. Lipsa sau inconsistența datelor sunt corectate cel mai bine prin consultarea respondenților în timpul culegerii datelor sau imediat după aceea.

Dat fiind impactul deosebit al operațiunilor de culegere și captare a datelor asupra calității datelor, este esențială utilizarea unor instrumente corespunzătoare de măsurare a calității și performanțelor pentru a administra aceste procese și a furniza măsuri obiective supervisorilor și clienților.

De-a lungul procesului trebuie urmăriți niște pași corespunzători pentru păstrarea confidențialității și securității informațiilor culese (chestiune discutată în Secțiunea 3.13).

Recomandări

- Anchetatorii și operatorii de captare a datelor sunt extrem de importanți pentru succesul majorității operațiunilor de culegere și captare a datelor. Asigurați-vă că aceștia beneficiază de o instruire și instrumente de lucru corespunzătoare.
- Valorificați avantajele utilizării tehnologiei informației și comunicațiilor pentru a îmbunătăți eficiența și eficacitatea culegerii și captării de date și pentru a reduce costurile și riscurile. Aveți în vedere interviuarea cu ajutorul unui computer (CAPI și CATI) și raportarea electronică a datelor (EDR) cu ajutorul internetului, introducerea automată a datelor (utilizând ICR) și codificarea automată prin recunoașterea textului (ACTR).
- Stabiliți proceduri corespunzătoare de control al eșantionelor pentru toate operațiunile de culegere a datelor. Instituiți sisteme de control eficiente pentru a vă asigura de protejarea captării, transmisiei și manipulării datelor. Urmăriți starea unităților eșantionate din momentul selectării eșantionului pentru a curăța datele editate astfel încât responsabilii de culegere a datelor și anchetatorii să poată evalua progresul oricând.
- Controlați operațiunile de livrare prin poștă a chestionarelor pe hârtie pentru a vă asigura că fiecare unitate selectată, primește chestionarul potrivit. Chestionarul o dată returnat, verificați precizia acoperirii informației și calitatea datelor furnizate. În unele cazuri pot fi necesare interviurile de recontactare. Când nu se primește nici un chestionar, sunt necesare activitățile de verificare pentru stabilirea stării unității (ex. activă sau inactivă) și pentru obținerea informației lipsă.
- Când culegeți date, asigurați-vă că persoana potrivită din cadrul gospodăriei sau organizației respondente este contactată la timpul potrivit, astfel încât informația să fie disponibilă. Permiteți-i respondentului să vă ofere datele într-un mod și un format convenabil. Acest lucru va ajuta la creșterea ratelor de răspuns și a calității datelor.
- Asigurați-vă că procedurile se aplică la fel tuturor unităților. Automatizarea este de preferat și permite personalului să lucreze la cazurile dificile. Centralizați prelucrarea pentru a reduce costurile și a beneficia de cunoștințelor experților. Având în vedere că se pot cere modificări în proceduri pe viitor, elaborați procese ce pot fi ușor de modificat.
- Monitorizați frecvența greșelilor de editare și a rectificărilor pe strat, mod de culegere, tip de prelucrare și tip de date. Acest lucru va ajuta la evaluarea calității datelor și la eficiența procesului de editare.
- Cuantificarea cheltuielilor, performanțelor și calității în timpul operațiunilor de culegere a datelor îi dă posibilitatea unui responsabil de anchetă să ia decizii cu privire la nevoia de modificare sau reconfigurare a procesului. Înregistrați costurile efective legate de trimiterea prin poștă, apelurile telefonice, elaborarea chestionarului, calculele și timpul consumat de fiecare persoană în fiecare zi. Măsurători calitative importante includ ratele de răspuns, procesarea ratelor de eroare, ratele de recontactare și numărarea non-răspunsurilor pe motive.
- Captarea manuală a datelor din chestionarele pe hârtie este supusă erorilor de codificare. Încorporați prelucrări on-line astfel încât operatorul responsabil cu captarea datelor să poată corecta erorile de codificare și să le înregistreze pentru revizuirile și analizele ulterioare.
- Folosiți metode statistice de control al calității pentru a evalua și îmbunătăți calitatea operațiunilor de culegere și captare, în special captarea și codificarea datelor. Culegeți și analizați rezultatele controlului calității pentru a identifica cauzele principale ale erorilor și raportați managerilor, personalului, specialiștilor în domeniu și metodologilor.
- Utilizați procese ulterioare de control al calității pentru a aduna informații utile în ceea ce privește calitatea, care pot semnală faptul că procedurile și instrumentele de culegere și captare ar putea necesita schimbări în cadrul ciclurilor viitoare de anchetă. De exemplu, editarea și evaluarea (vezi Secțiunile 3.8 și 3.12) pot sugera probleme legate de culegere.

Referințe

Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L. and O'Reilly, J. (eds.) (1998). *Computer Assisted Survey Information Collection*. Wiley, New York.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. Wiley, New York.

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. and Trewin, D. (eds.) (1997). *Survey Measurement and Process Quality*, Wiley, New York.

Mudryk, W. and Xie, H. (2002). *Quality control application in ICR data capture for the 2001 Census of Agriculture*. Proceedings of the Section on Quality and Productivity, American Statistical Association, 2424-2429.

Mudryk, W., Burgess, M.J. and Xiao, P. (1996). *Quality control of CATI operations at Statistics Canada*. Proceedings of the Section on Survey Research Methods, American Statistical Association, 150-159.

3.7 Non-răspunsul

Scop și obiective

În ciuda celor mai susținute eforturi ale personalului de anchetă de a maximiza răspunsul, vor apărea totuși non-răspunsuri. Non-răspunsul poate fi de două tipuri. *Non-răspunsul pe variabilă* este atunci când o anumită variabilă dintr-un chestionar nu a fost completată; iar *non-răspunsul pe unitate* este atunci când nu a fost primit nici un răspuns din partea unității sau răspunsul este atât de limitat, încât este efectiv inexistent.

Cele două abordări privind gestionarea datelor lipsă sunt imputarea (vezi Secțiunea 3.9) și ajustarea ponderii eșantionului (Secțiunea 3.10). În mod specific, ponderile sunt ajustate pentru non-răspunsul pe unitate și imputarea este utilizată pentru non-răspunsul pe variabilă.

Non-răspunsul are două efecte asupra datelor. Primul introduce abaterea în estimări când non-respondenții diferă de respondenți, iar metodele de corecție sunt aceleași. Al doilea crește dispersia eșantionului, deoarece dimensiunea efectivă a eșantionului este redusă.

Principii

Programul eficient de stabilire a relațiilor cu respondenții și chestionarul bine proiectat sunt elemente critice în maximizarea răspunsului.

Non-răspunsul este monitorizat pentru a fi transmis ca feed-back personalului de anchetă, în vederea acțiunii imediate și pe viitor și se raportează utilizatorilor.

Eforturi pentru a obține un răspuns în orice caz dat se bazează pe considerații privind bugetul, oportunitatea și impactul asupra calității datorată abaterii cauzate de non-răspuns. Analizele ce trebuie inițiate pentru dezvoltarea tehnicilor de ajustare a non-răspunsului sunt de asemenea influențate de aspectele legate de buget, timp, utilizarea datelor și riscul de abatere.

Recomandări

- Rata de răspuns mare necesită o abordare integrată, cuprinzând următoarele:
 - Unități bine definite ale cadrului;
 - Un cadru actualizat al anchetei în termenii datelor de acoperire și contact;
 - Perioadă de referință bine definită și sincronizată cu perioada de culegere;
 - Metodele potrivite de culegere a datelor; sarcină minimă de răspuns (în ceea ce privește lungimea chestionarului, dificultatea întrebărilor, frecvența culegerii); atenție la domeniile "sensibile" (sensibilitatea subiecților);

- Metodologia de urmărire eficientă și anticipare a dificultăților în depistarea respondenților care s-au mutat – utilizarea pre-testării și/sau rezultatele din anchetele similare pentru a stabili ratele de răspuns așteptate;
- Utilizarea eficientă a stimulentelelor pentru respondenți;
- Experiență anterioară cu același tip de anchetă și abilitate demonstrată a personalului care se ocupă de culegerea datelor;
- Sarcină de lucru rezonabilă a personalului care se ocupă de culegerea datelor;
- Cunoaștere aprofundată a respondenților și relațiile cu aceștia;
- Buget suficient și alocarea adecvată pentru diverse operațiuni;
- Recontactarea non-respondenților în limita bugetului. Recontactarea crește rata de răspuns și poate ajuta la determinarea similitudinilor în ceea ce privește caracteristicile măsurate ale respondenților și non-respondenților.
- Recontactarea este în mod special importantă în cazul anchetelor longitudinale în care eșantionul este supus degradării progresive din cauza non-răspunsului cu ocazia fiecărei anchete. Facilitați detectarea de înaltă calitate utilizând datele administrative, cărțile de telefoane și alte surse locale de informații. Verificați/actualizați informațiile de contact cu ocazia fiecărei anchete și solicitați unităților să notifice INS în privința oricăror schimbări ce pot apărea de la un ciclu al anchetei la altul.
- Prioritizați activitățile de recontactare. De exemplu, în anchetele în întreprinderi, recontactarea mai întâi a unităților mari și a celor cu influență. Tot astfel, puneți ca prioritate unitățile non-respondente în domenii cu potențial ridicat de abatere datorată non-răspunsului. Utilizați un sistem de punctare pentru prioritizarea recontactărilor.
- Înregistrați și monitorizați motivele pentru non-răspuns (ex.: refuz, necontactare, absent temporar, probleme tehnice). Gradul abaterii datorată non-răspunsului poate diferi în funcție de motiv. Monitorizați tendințele de non-răspuns pe motive.
- Încercați să determinați diferențele (dacă există) între respondenți și non-respondenți prin legătura cu sursele externe de date (de exemplu, fișierele de date administrative) și prin examinarea răspunsurilor non-respondenților care au devenit respondenți în timpul recontactării. În cazul anchetelor periodice, caracteristicile cunoscute ale respondenților pentru un ciclu/val al anchetei pot fi folosite pentru compararea caracteristicilor respondenților și non-respondenților din ciclurile/valurile ulterioare.
- Încercați să evaluați măsura în care procedurile de ajustare a non-răspunsului corectează abaterea potențială. Luați în considerare non-răspunsul atunci când produceți estimări și estimarea dispersiei asociate.
- Înregistrați ratele ponderate și neponderate ale non-răspunsului în etapa estimării.
- Informați utilizatorii despre rata de non-răspuns atunci când furnizați estimări.

Referințe

Cialdini, R., Couper, M. and Groves, R.M. (1992). *Understanding the decision to participate in a survey*. Public Opinion Quarterly, 56, 475-495.

Federal Committee on Statistical Methodology (2001). *Measuring and reporting sources of error in surveys*. Statistical Policy Working Paper 31. See also <http://www.fcsm.gov>.

Groves, R.M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (2002). *Survey Non-response*. Wiley, New York.

3.8 Editarea

Scop și obiective

Editarea datelor reprezintă efectuarea unor verificări pentru a detecta datele lipsă, neadecvate sau fără coerență sau de a indica înregistrările de date care pot conține eventuale erori. Unele din aceste verificări se bazează pe relații logice care reies direct din definițiile și relațiile dintre date. Altele au o natură mai empirică – verificările fezabilității, tehnici de detectare a valorilor aberante. Verificările se pot baza pe date culese din ciclurile anterioare ale aceleiași anchete sau din alte surse.

Editarea presupune o mare varietate de activități, de la verificări pe teren ale anchetatorilor, avertismente electronice din timpul culegerii sau captării de date, identificarea unor unități pentru recontactare, localizarea erorilor în scopul imputării, până la validarea datelor. Ultimele două subiecte vor fi discutate în Secțiunile 3.9 și 3.12.

Principii

Editarea are trei scopuri:

- să corecteze datele;
- să ofere informații legate de calitatea datelor;
- să furnizeze baza pentru îmbunătățirea viitoare a chestionarului și operațiunilor anchetei.

Evitați să consumați o cantitate disproporționată de resurse ale anchetei pentru corectare datelor.

Erorile fatale (intrările nevalide sau inconsistente) ar trebui scoase din date, pentru a facilita analiza și prelucrarea automată ulterioară a datelor și pentru a asigura credibilitatea INS. Identificarea erorilor fatale poate și trebuie să fie automată. Totuși există pericolul utilizării peste măsură a *interogărilor de editare* (cele care indică intrări sau înregistrări care au potențialul de a fi eronate), mai ales din cauza faptului că editarea este, probabil, cea mai scumpă activitate a unei anchete, valoarea ei ridicându-se poate la un sfert din bugetul total.

Ratele de performanță a editării, definite ca proporția de editări care descoperă erori reale sunt, adesea, destul de joase, de exemplu 20-30%. Mai mult, nu este ieșit din comun ca cele câteva corecții ale erorilor să fie responsabile pentru schimbările majore ale estimărilor datorită editării. Când impactul majorității interogărilor de editare asupra estimărilor finale este neglijabil, se spune că a apărut supra-editarea.

Editarea joacă un rol important în furnizarea de informații pentru îmbunătățirea ulterioară a procesului anchetei - în cizelarea definițiilor, îmbunătățirea chestionarului, evaluarea calității datelor, identificarea surselor de eroare care nu provin din eșantionare.

Recomandări

- Asigurați-vă că regulile de editare au o consecvență internă.
- Supuneți unitățile unui proces de reeditare în cazul efectuării rectificărilor pentru a vă asigura că nu au mai fost introduse erori în urma acestui proces.
- Editați în mod selectiv. Prioritățile se pot stabili în funcție de tipurile și/sau gravitatea erorii și/sau în funcție de importanța datelor și/sau a unităților de raportare.
- În cazul anchetelor în întreprinderi, concepeți o strategie pentru recontactarea selectivă. Folosirea unui sistem de punctare este un mijloc de a asigura că resursele sunt concentrate sistematic pe cele mai importante unități, pe indicatorii cheie și pe erorile severe.
- Aveți în vedere reducerea efortului de editare și redirecționarea resurselor în activități cu o rentabilitate mai mare, de exemplu, analiza datelor și analiza erorilor de răspuns.

- Acordați o prioritate ridicată învățării din procesul editării. Pentru a reduce erorile, practicați mai degrabă prevenirea erorilor, decât corecția lor. În acest scop, asigurați-vă că editarea are loc cât mai devreme posibil în procesul anchetei, preferabil cât timp respondentul este încă disponibil, de exemplu, cu ajutorul unor metode telefonice asistate de computer, metode personale sau de auto-intervievare.
- Editările nu pot, sub nicio formă, să identifice erorile sistematice minore, înregistrate în mod constant în anchetele periodice. Editările stricte nu sunt o soluție. Aceste erori sunt cel mai bine detectate și corectate prin analiza și revizuirea conceptelor și definițiilor, prin studiile post-interviu, prin validarea datelor și confruntarea datelor cu alte surse de date.
- Identificați aberațiile (valorile extreme ale datelor) în perioada unei anchete sau de-a lungul perioadelor de desfășurare a cercetării. Acestea sunt avertismente asupra erorilor potențiale.
- Atunci când realizați recontactarea, nu supraestimați capacitatea respondenților de a raporta sau de a corecta raportările anterioare. Memoria lor este limitată și pot furniza date diferite.
- Procesul de editare este, de multe ori, foarte complex. Asigurați-vă că procedurile sunt descrise corespunzător și actualizate, oferiți o instruire corespunzătoare personalului implicat și monitorizați munca în sine. Aveți în vedere folosirea unor proceduri formale de control al calității.
- Automatizați pe cât posibil editarea. Deși pot fi necesare unele intervenții manuale, software-ul generalizat, reutilizat vine în sprijinul acestui scop. Totuși, nu fiți tentați să creșteți scopul și volumul verificărilor doar pentru că acestea se pot face automatizat. Controalele prin interogări, care produc mici diferențe în estimări, trebuie excluse.

Referințe

Bankier, M., Lachance, M. and Poirier, P. (1999), *A generic implementation of the New Imputation Methodology*. Proceedings of the Survey Research Methods Section, American Statistical Association, 548-553.

De Waal, T., Van de Pol, F. and Renssen, R. (2000). *Graphical macro editing: possibilities and pitfalls*. Proceedings of the Second International Conference on Establishment Surveys, Buffalo, New York.

Granquist, L. and Kovar, J.G. (1997). *Editing of survey data: how much is enough?* In Survey Measurement and Process Quality, Lyberg et al (eds.). Wiley, New York, 415-435.

Latouche, M. and Berthelot, J.-M. (1992). *Use of a score function to prioritize and limit recontacts in editing business surveys*. Journal of Official Statistics, 8, 389-400.

3.9 Imputarea

Scop și obiective

Imputarea este procesul de înlocuire a valorilor lipsă sau nevalide sau inconsistente, creând astfel înregistrări solide, și în același timp păstrând cât mai multe din valorile originale ale datelor.

Ideal, intrările lipsă, nevalide și inconsistente ar trebui rezolvate în stadiul de culegere a datelor sau în etapa de recontactare. Totuși, în practică acest lucru este imposibil din cauza sarcinii de răspuns, a constrângerilor legate de costuri și de timp.

Unele metode de imputare pot deforma distribuțiile de frecvență ale datelor de bază și relațiile dintre acestea. Acest lucru trebuie luat în calcul când se produc estimări și estimările dispersiei aferente.

Principii

Imputarea poate fi automată, manuală sau o combinație a acestora.

Procesele unei bune imputări:

- Depind de accesul la micro-date și la o bună informație auxiliară;
- Se asigură că înregistrările imputate sunt consistente;
- Generează un număr minim de modificări ale datelor dintr-o înregistrare;
- Limitează abaterea cauzată de neobservarea tuturor valorilor dorite;
- Sunt automate, obiective, reproductibile și eficiente;
- Creează posibilitatea realizării unui audit în scopuri de evaluare.

Metodele de imputare pot fi clasificate ca deterministe și stocastice.

- Metodele de *imputare deterministă* includ imputarea logică, imputarea istorică, imputarea medie, imputarea de proporție și regresie și imputarea prin cel mai apropiat donor vecin. Aceste metode pot fi împărțite mai departe în metode care se sprijină doar pe deducerea valorii imputate din datele disponibile pentru non-respondent și/sau din alte date auxiliare despre respondent și acelea care utilizează datele observate pentru alte unități respondente din acea anchetă.
- Metodele de *imputare stocastică* includ hot deck-ul, imputarea de la cel mai apropiat vecin unde selecția aleatoare se face de la mai mulți vecini din proximitate, regresia cu reziduuri aleatorii și orice altă metodă deterministă cu reziduuri aleatorii adăugate.

Recomandări

- Determinați care dintre datele auxiliare pot explica cel mai bine mecanismele de non-răspuns și utilizați-le spre perfecționarea metodei de imputare. Studiați calitatea și adecvarea datelor disponibile și selectați-le pe cele mai potrivite și pe care să le folosiți la construirea claselor de imputare.
- Luați în considerare tipurile de estimări ce trebuie produse, pe nivel sau schimbare, pe agregate mari sau domenii restrânse și pe date transversale sau longitudinale.
- Adesea este adecvată utilizarea mai multor tipuri diferite de metode de imputare într-o succesiune ordonată.
- Asigurați-vă că înregistrările imputate sunt consistente pe plan intern.
- Asigurați-vă că înregistrările imputate seamănă cu cele originale cât mai mult posibil.
- Testați metoda de imputare. Consultați experții în acest domeniu.
- Când se utilizează imputarea prin donor, încercați să imputați datele pentru o înregistrare de la cât mai puțini donori posibili. Operațional, acest lucru înseamnă adesea un donor pe fiecare secțiune a chestionarului. De asemenea, acordați donatorilor buni o probabilitate egală de a fi selectați și evitați sistematic selectarea unui donor.
- Pentru anchetele mai ample, poate fi mai degrabă necesară imputarea datelor în două sau mai multe etape, decât într-una singură, astfel încât să se reducă costurile de calculare sau pentru că există multe câmpuri lipsă sau incorecte în înregistrări.
- Marcați valorile imputate și identificați cu claritate metodele și sursele de imputare. Rețineți valorile anterioare și pe cele imputate pentru evaluări.
- Evaluați gradul și efectele imputării, în special modalitatea de a măsura dispersia de eşantionare sub imputare.
- Aveți în vedere gradul și impactul imputării atunci când analizați datele. Imputarea ar putea avea un impact semnificativ asupra distribuției datelor. De pildă, valorile dintr-un domeniu au crescut sistematic, pe când valorile dintr-un alt domeniu au scăzut printr-o cantitate compensatorie, lăsând totalurile la fel. Cu cât gradul și impactul imputării este mai mare, cu

atât mai rațional trebuie să fie analistul atunci când utilizează datele. În unele cazuri, analizele pot fi mai bine elaborate, dacă se exclud valorile imputate.

Software

- Pe durata dezvoltării metodologiei de imputare, utilizați oricare dintre sistemele de imputare generalizată care sunt disponibile. La INS, estimarea cantitativă a datelor este în prezent realizată prin utilizarea macrouilor SAS mainmean.sas și a macmean.sas și a variabilelor ordinale și nominale prin macroui SAS: ini_execute.sas and execute.sas.

Referințe

Bankier, M., Lachance, M. and Poirier, P. (1999). *A generic implementation of the New Imputation Methodology*. Proceedings of the Survey Research Methods Section, American Statistical Association, 548-553.

Fellegi, I.P. and Holt, D. (1976). *A systematic approach to automatic edit and imputation*. Journal of the American Statistical Association, 71, 17-35.

Kovar, J.G., and Whitridge, P. (1995). *Imputation of business survey data*. In Business Survey Methods, B.G. Cox et al. (eds.), Wiley, New York, 403-423.

Lee, H., Rancourt, E. and Särndal, C.-E. (2002). *Variance estimation from survey data under single imputation*. In Survey Non-response, R.M. Groves et al. (eds.), Wiley, New York, 315-328.

3.10 Estimarea

Scop și obiective

Estimarea este procesul care generează estimări ale parametrilor populației pe baza datelor din eșantion. Exemple de parametri includ statistici simple descriptive precum totalurile, mediile, proporțiile și procentajele, precum și statistici analitice mai complicate precum coeficienții de regresie. Inferența poate fi sub forma estimărilor punctuale, intervalelor de confidență, rezultatelor testelor ipotezei. Estimările punctuale sunt de obicei însoțite de estimări ale preciziei.

Acolo unde parametrii populației sunt funcții specifice ale totalurilor populației, estimatorii acestora sunt în general funcții corespondente ale totalurilor estimate ale populației.

Se pot folosi date *auxiliare* pentru a îmbunătăți precizia estimațiilor.

Principii

Eroarea totală în anchetă a unei estimări este gradul prin care diferă de valoarea adevărată a parametrului populației care a fost estimat. Poate fi considerată ca suma *erorii de eșantionare* și a *erorii de non-eșantionare*. Eroarea de eșantionare reprezintă eroarea asociată cu estimarea pe date doar din eșantion. Aceasta include abaterea eșantionării și dispersia de eșantionare. Erorile de non-eșantionare includ toate celelalte cauze ale lipsei de precizie, incluzând erorile de acoperire, erorile de măsurare și erorile de non-răspuns, după cum se va discuta în alte secțiuni.

Metoda de estimare și proiectarea eșantionului determină ambele proprietăți ale erorii de eșantionare. Obiectivul este de a alege un estimator (adică o metodă de estimație) care conferă o abatere și o dispersie minimă și este consistent, adică eroarea de eșantionare este zero când eșantionul cuprinde întreaga populație.

Când nu sunt disponibile informații auxiliare utile pentru estimare, cea mai naturală alegere a estimatorului este *estimatorul Horvitz-Thompson*. El atribuie unităților ponderi care reprezintă inversul probabilităților de includere ale unităților respective. Aceste ponderi se numesc *pondere de eșantionare*. Ponderea de eșantionare poate fi interpretată ca numărul care arată de câte ori trebuie replicată o unitate pentru a reprezenta întreaga populație.

Estimatorul poate fi îmbunătățit atunci când sunt disponibile date auxiliare. *Calibrarea* este o procedură prin care se pot încorpora datele auxiliare. Această procedură ajustează ponderile de eșantionare prin intermediul multiplicatorilor cunoscuți ca *factori de calibrare* care pun de acord estimările cu totalurile cunoscute. Ponderile care rezultă din aceasta se numesc *ponderi de calibrare* sau *ponderi de estimare finală*. Aceste ponderi de calibrare vor rezulta în general în estimări consistente care au o dispersie de eșantionare mai mică decât estimatorul Horvitz-Thompson. *Metodele de delimitare* sunt folosite pentru a lua măsuri împotriva producerii de ponderi negative și foarte mari în urma calibrării.

Ca rezultat al non-răspunsului, eșantionul observat are dimensiuni mai mici decât eșantionul original selectat. Pentru a compensa un non-răspuns, se folosește de obicei *reponderarea*. Reponderarea constă în ajustarea ponderilor de eșantionare prin *factorii de ajustare a non-răspunsurilor* înainte de calibrare. Principiul de bază în calcularea factorilor de ajustare a non-răspunsurilor este utilizarea inversului probabilității de răspuns. Totuși, acestea sunt necunoscute și trebuie estimate. Soluția pentru reducerea abaterilor non-răspunsului este să se obțină un model de non-răspuns profitând cât mai mult posibil de datele auxiliare disponibile.

Valorile aberante pot conduce la estimări lipsite de fiabilitate pentru variabilele continue. *Valorile aberante* ar putea să se datoreze fie valorilor extreme ale unor date, fie ponderilor foarte mari, sau ambelor. Este necesară o anumită formă de detectare și ajustare a valorilor aberante.

În cadrul anchetelor longitudinale, derivă de obicei două seturi de ponderi de estimare: *ponderile longitudinale* și *ponderile transversale*. Ponderile longitudinale se referă la populația din selecția inițială a eșantionului longitudinal. Aceste ponderi sunt de obicei ajustate pentru a lua în considerare deteriorarea eșantionului în timp. Ele sunt folosite atunci când se efectuează analiza datelor longitudinale. Ponderile transversale sunt legate de populația stabilă pentru fiecare val al anchetei. Aceste ponderi sunt în mod normal utilizate pentru a produce estimări punctuale sau diferențe ale estimărilor punctuale între două perioade de timp. Din cauza schimbărilor populației de-a lungul timpului, cele două seturi de ponderi sunt în general diferite.

În anchetele periodice sau longitudinale cu un grad mare de suprapunere a eșantionului între perioade, aveți în vedere utilizarea metodelor de estimare care exploatează corelația în timp. În esență, aceste metode tratează datele din momentele anterioare, ca variabile auxiliare. O astfel de metodă este denumită *estimarea compusă*.

Recomandări

- Încorporați ponderile de eșantionare în procesul de estimare. Aceasta implică reflectarea tuturor aspectelor relevante ale proiectării eșantionului - stratificarea, aglomerarea (clustering) și informațiile multifazice sau în mai multe etape - în procedurile de estimare.
- Când sunt disponibile date auxiliare pentru unitățile de eșantionare împreună cu totalul populației cunoscute pentru astfel de date, aveți în vedere calibrarea pentru ca prin datele auxiliare ponderate însumate să se obțină aceste totaluri cunoscute. Aceasta poate determina un grad mai mare de precizie și coerență între estimări din mai multe surse.
- În cazul în care clasificarea inițială a unităților de eșantionare s-a modificat de la data selecției eșantioanelor și până la data estimății, luați în considerare *estimația domeniului*, pentru ca noua clasificare să se reflecte în estimări. Estimația domeniului se referă la estimația pentru subseturi specificate ale populației (sau *domeniilor*) de interes. Adesea, unitățile din aceste subseturi nu au fost sau nu au putut fi identificate înainte de eșantionare. Estimarea în prezența unor unități de eșantion decedate sau dinafara domeniului reprezintă un exemplu de estimație a domeniului.
- Calitatea factorilor de ajustare a non-răspunsurilor depinde de ipotezele făcute în modelul (explicit sau implicit) non-răspunsului. Modelul trebuie validat prin utilizarea datelor auxiliare corelate cu probabilitatea de a răspunde. Pentru a vă proteja împotriva eșecului datorat modelului, formați clase de ajustare pentru non-răspunsuri și estimați probabilitățile de răspuns prin ratele de răspuns din cadrul acestor clase.

- Teoria de eșantionare în două faze poate fi folosită pentru a estima dispersia pentru mai mulți estimatori, încorporând ajustările de non-răspuns.
- Atunci când e cazul, folosiți eșantionarea în două faze pentru a îmbunătăți estimarea prin încorporarea datelor auxiliare. Datele auxiliare pot fi folosite (a) pentru a stratifica eșantionul din faza a doua, (b) pentru a îmbunătăți estimația prin intermediul estimatorului de diferență, proporție sau regresie, sau (c) pentru a extrage un subeșantion de unități non-respondente. O abordare generală a eșantionării în două faze este atunci când datele auxiliare sunt încorporate în procesul de estimare prin Estimatorul Generalizat de Regresie.
- Cerințele pentru domenii restrânse, precum zonele mici, sunt cel mai bine încorporate în etapa proiectării eșantionului. Totuși, dacă acest lucru nu este posibil în etapa de proiectare sau dacă domeniile sunt abia mai târziu specificate, aveți în vedere metode speciale de estimare în etapa estimației. Aceste metode împrumută forța din domeniile asociate pentru a minimiza eroarea medie pătratică a estimatorului rezultat. Estimatorii pentru ariile mici sunt unele dintre exemple.
- Pentru a trata valorile aberante, utilizați estimatorii rezistenți la valori aberante (robuști).
- Ori de câte ori este posibil, utilizați de preferință software-ul de estimare generalizată decât sistemele personalizate. Posibile pachete software: SAS, FOX, SUDAAN, PC CARP, STATA.

Referințe

Chambers, R.L., Kocic P., Smith P. and Crudas M. (2000). *Winsorization for identifying and treating outliers in business surveys*. Proceedings of the Second International Conference on Establishment Surveys, June 17-21, 2000, Buffalo, New York, 717-726.

Cochran, W.G. (1977). *Sampling Techniques*. Wiley, New York.

Deville, J.-C. and Särndal, C.E. (1992). *Calibration estimators in survey sampling*. Journal of the American Statistical Association, 87, 376-382.

Fuller, W.A. (2002). *Regression estimation for survey samples*. Survey Methodology, 28, 5-23.

Hidiroglou, M.A. and Särndal, C.E. (1998). *Use of auxiliary information for two-phase sampling*. Survey Methodology, 24, 11-20.

Lee, H. (1995). *Outliers in business surveys*. In Business Survey Methods, B.G. Cox et al. (eds). Wiley, New York, 503-526.

Rao, J.N.K. (1999). *Some recent advances in model-based small area estimation*. Survey Methodology, 25, 175-186.

Singh, A.C., Kennedy, B. and Wu, S. (2001). *Regression composite estimation for the Canadian Labour Force Survey with a rotating design*. Survey Methodology, 23, 33-44.

Singh, M.P., Hidiroglou, M.A., Gambino, J. and Kovacevic, M. (2001). *Estimation methods and related systems at Statistics Canada*. International Statistical Review, 69, 461-486.

Thompson, M.E. (1997). *Theory of Sample Surveys*. Chapman and Hall.

3.11 Ajustarea sezonieră și estimarea ciclului de trend

Scop și obiective

Seriile macroeconomice infra-aniuale reprezintă cheia elaborării politicii economice. Totuși, acestea sunt influențate de fluctuațiile sezoniere și efectele zilelor calendaristice/zile lucrătoare, care maschează mișcările relevante pe termen scurt și lung și împiedică înțelegerea clară a fenomenelor economice fundamentale. Principalul obiectiv al ajustării sezoniere este de a filtra fluctuațiile sezoniere obișnuite și efectele tipic calendaristice.

O serie cronologică este văzută ca având patru componente – o tendință, un efect ciclic, un efect sezonier și o componentă neregulată.

- *Tendința* este mișcarea de bază pe termen lung care durează mai mulți ani.
- *Ciclul*, numit de obicei ciclul de afaceri, este oscilația cvasiperiodică ce durează câțiva ani.
- Tendința și ciclul sunt de obicei estimate în combinație.
- *Componenta neregulată* reprezintă acele variații aleatorii care sunt mișcări neprevăzute, legate de diferite evenimente.

Ajustarea sezonieră constă în estimarea factorilor sezonieri și aplicarea lor la o serie cronologică pentru a înlătura *variațiile sezoniere*. Aceste variații reprezintă efectul compus al factorilor climatici și instituționali care se repetă cu o anumită regularitate într-un an.

O serie ajustată sezonier constă din ciclul de trend și o componenta neregulată. Omogenizarea seriilor ajustate sezonier în vederea eliminării componentei neregulate produce *estimările privind ciclul de trend*, care furnizează imaginea ciclului de afaceri de bază – direcția mișcării, amplitudinea ciclului și, mai ales, coordonarea punctelor de inflexiune.

În contextul ajustării sezoniere, scopul *pre-tratării* este de a asigura estimarea fiabilă a componentei sezoniere prin detectarea și corectarea seriei pentru probleme specifice cunoscute înainte de estimare. Pre-tratarea include analiza grafică, ajustările calendaristice (pentru zilele lucrătoare, vacanțe și calendarul UE), detectarea și corectarea valorilor aberante, tratarea datelor lipsă și selecția *modelului și schemei dezagregării*.

Principii

O serie cronologică este ajustată sezonier numai atunci când există dovezi că seria este influențată de forțe sezoniere și că seria conține sezonalitate identificabilă. *Sezonalitatea identificabilă* este definită ca model sezonier care nu este ascuns printr-un grad înalt de fluctuații neregulate și astfel poate fi identificat în mod sigur.

În cel mai bun caz, procedura de ajustare sezonieră nu lasă sezonalitate reziduală în serie, iar seria ajustată sezonier care rezultă este întrucâtva mai omogenă decât cea inițială. Seria ciclului de trend este mult mai omogenă, deoarece s-a eliminat componenta neregulată.

Ultimele câteva puncte din seria ajustată sezonier sunt cele mai predispuse la eroare și deci și la revizie, pe măsură ce mai multe puncte devin disponibile. Cel mai bine ar fi ca reviziile estimărilor ajustate sezonier să fie minime.

Pentru seriile agregate ce rezultă dintr-o combinație a unor serii componente, acele serii componente care conțin sezonalitate identificabilă trebuie să fie ajustate sezonier, iar celelalte trebuie lăsate neajustate. Seria agregată însăși trebuie să fie ajustată astfel încât să nu conțină sezonalitatea reziduală și să fie relativ omogenă. Poate fi ajustată prin metoda indirectă sau directă.

- În *metoda indirectă*, componentele ajustate sezonier și cele neajustate se combină pentru a obține agregatul ajustat sezonier.
- În *metoda directă*, componenta agregată este ajustată sezonier și caracteristica sa aditivă (fiind suma componentelor) este restabilită prin înclinarea/calibrarea componentelor ajustate sezonier.

Ghidul SSE privind ajustarea sezonieră

Ghidul SSE publicat în 2008 asigură o descriere definitivă și cuprinzătoare a tuturor aspectelor legate de ajustarea sezonieră. Acestea se bazează pe două abordări:

- TRAMO-SEATS - sprijinită de Banca Spaniei
- X-12-ARIMA - sprijinită de Biroul de Recensăminte ale SUA

Acestea acoperă pre-tratarea seriei, politicile de revizuire, calitatea și documentarea. Ele pun în discuție limitările ajustării sezoniere, ex. serie pe termen scurt și asigură trei seturi de alternative: a) cea mai bună abordare; b) abordare acceptabilă; c) practici ce trebuie evitate

Recomandări

- Folosiți publicația cea mai recentă a *Ghidului SSE privind ajustarea sezonieră*.
- Alegerea unei metode dintre cele două recomandate de ajustarea sezonieră ține de experiența, subiectivitatea și caracteristicile seriei de timp.
- Înainte de a ajusta sezonier pentru prima dată o serie de timp, elaborați o analiză minuțioasă pentru a stabili dacă sezonalitatea este identificabilă. Publicați seriile în forma ajustată sezonier numai dacă este identificabilă. Verificați situația, ulterior, o dată la câțiva ani.
- Frecvența reviziilor trebuie minimizată. Deși apar revizii cu fiecare punct de date adăugat, implementați revizuirii numai atunci când acestea determină îmbunătățirea estimărilor, adică atunci când estimarea revizuită devine în mod semnificativ mai apropiată de estimarea finală.
- Ori de câte ori sunt publicate cifre ajustate sezonier care aparțin aceleiași activități economice, coordonați opțiunile de ajustare sezonieră aplicate pe ariile implicate și faceți toate eforturile să tratați seriile relaționate într-un mod consecvent.
- Revizuiți anual valorile ajustate sezonier pentru ultimii trei ani atunci când datele pentru prima perioadă (lună, trimestru) din anul următor devin disponibile. Atunci când valorile ajustate sezonier sunt obținute prin factori sezonieri previzionați (prognozați), revizuirea anuală trebuie să se aplice la ultimii patru ani.
- Pentru seriile agregate ce cuprind combinațiile seriilor componente, alegeți abordarea directă, dar numai dacă calibrarea/înclinarea nu deformează prea mult modificările de la lună la lună (de la trimestru la trimestru) ale seriilor componente.
- De fiecare dată când sunt publicate cifre ajustate sezonier aparținând aceleiași activități economice, coordonați opțiunile de ajustare sezonieră aplicate la ariile implicate. De exemplu, ori de câte ori este posibil, alegeți între ajustarea directă și cea indirectă a seriilor compuse și asigurați-vă că evenimentele extreme din seriile temporale cauzate de fapte precum închiderea fabricilor, greve, calamități naturale, etc. sunt tratate de manieră coerentă pe fiecare arie în parte.
- Pentru ajustarea sezonieră a observațiilor recente, folosiți un *factor sezonier concurrent*. Acesta este un factor obținut prin utilizarea tuturor punctelor de date disponibile. Se poate aplica o excepție la această recomandare, și anume atunci când cele mai recente observații au făcut obiectul unor mari schimbări. În acest caz pot fi mai adecvați *factorii sezonieri din anii următori (previziune)* pe baza datelor înregistrate la finele anului precedent.

Când se folosește Estimația ciclului de trend X-12-ARIMA

- Folosiți ponderile zilnice care sunt estimate automat pentru serii cu variații ale zilelor lucrătoare. În cursul anului curent păstrați-le fixe, furnizându-le drept ponderi zilnice prealabile. Vor fi modificate la următoarea revizuire anuală. Excepții la această recomandare pot apărea când sunt furnizate ponderi zilnice a-priori de către experți în materie, pe baza unei cunoașteri mai bune a seriilor în cauză.
- Folosiți factorii de efect Easter ("Paște") calculați automat de program.
- Utilizați subrutina automată de extrapolare ARIMA ori de câte ori este posibil. Dacă nu se selectează niciunul dintre modelele încorporate, atunci trebuie furnizat un model ARIMA.
- Folosiți mediile mobile Henderson pentru a produce estimări ale ciclului de tendință. Pentru a vă asigura că linia de tendință se află în interiorul diagramei de dispersie a seriilor ajustate sezonier, aplicați mediile mobile Henderson la seriile ajustate sezonier publicate.

- Înainte de aplicarea estimatorului ciclu de trend, extindeți seriile ajustate sezonier cu un an de valori anticipate dintr-un model ARIMA corespunzător seriilor ajustate sezonier.
- Aplicați mediile mobile Henderson seriilor ajustate sezonier extinse ale căror extreme au fost corectate anterior.
- Folosiți media mobilă Henderson selectată automat de programul X12-ARIMA: pe baza valorii neregulate/ a raportului ciclului de trend.
- Asigurați-vă că estimările ciclului de trend sunt compatibile cu estimările ajustate sezonier publicate.
- Informații utilizatorii că ultimele estimări ale ciclului de trend (mai ales ultimele) fac obiectul unor ample revizii. Atunci când un punct de date se adaugă la o serie poate avea loc chiar o inversare de mișcare. Incertitudinea asociată cu estimările de la sfârșit poate fi indicată de exemplu printr-o linie punctată pe graficul de tendință sau printr-un avertisment scris către utilizatori.

Referințe

Ghidul SSE privind Ajustarea Sezonieră (2009).

3.12 Evaluarea calității datelor

Scop și obiective

Evaluarea calității datelor este un proces care determină dacă rezultatele finale îndeplinesc obiectivele procesului statistic/anchetei. Rezultatele evaluării permit utilizatorilor să interpreteze mai bine rezultatele anchetei, iar Institutului Național de Statistică să își îmbunătățească calitatea ciclurilor curente și/sau viitoare ale anchetelor.

Există două metode principale de evaluare a calității datelor:

- *Certificarea* sau *validarea* este procesul în cadrul căruia datele sunt analizate înainte de a fi emise în mod oficial, în scopul de a se evita apariția erorilor grosiere și de a se elimina calitatea evident scăzută a unor informații. Acest proces coincide în mod frecvent cu o analiză interpretativă a datelor și este supus invariabil anumitor constrângeri de timp, putând astfel să implice doar metodele care produc rezultate rapide.
- *Sursele de studiere a erorilor* furnizează informații cantitative privind sursele specifice de erori apărute în date. În mod specific, rezultatele acestor studii sunt disponibile doar după publicarea oficială a datelor.

Principii

Utilizatorii trebuie să fie capabili să determine în ce măsură erorile apărute afectează utilizarea datelor. Totuși, întrucât aceștia rareori reușesc să evalueze în mod independent acuratețea, rămâne astfel la latitudinea fiecărei agenții să evalueze calitatea datelor și să ofere rapid utilizatorilor, într-un mod accesibil, rezultatele obținute.

Evaluările calității datelor sunt de asemenea utile pentru INS. În măsura în care erorile pot fi asociate cu anumite etape ale procesului de anchetă, evaluările pot conduce la îmbunătățirea calității în următorul ciclu de anchetă și/sau în alte anchete similare.

Sursele sau erorile care vor trebui evaluate cuprind următoarele:

- *Erori de acoperire*, care constau în omisiuni, includeri eronate și duplicări în cadrul folosit pentru realizarea anchetei. Acestea se pot traduce printr-o polarizare (bias) negativă sau pozitivă a datelor, iar impactul poate să varieze în funcție de populația anchetată. Erorile apărute în codificare, în special atunci când se aplică clasificări industriale și geografice pot duce la erori de acoperire.

- *Erori de eșantionare*, apar atunci când rezultatele anchetei sunt obținute dintr-un eșantion, și nu din anchetarea întregii populații. Acestea pot cuprinde de asemenea și *erori de estimare* datorită estimatorilor care, din rațiuni de proiectare sau din alte cauze, provoacă abateri, ca de exemplu în cazul estimatorilor pe arii mici.
- *Erori de non-răspuns*, apar atunci când nu există niciun răspuns la una sau la toate întrebările din cadrul anchetei. Non-răspunsul conduce la o creștere a gradului de dispersie ca urmare a reducerii dimensiunii efective a eșantionului și ca rezultat al recurgerii la procedee de imputare, acest lucru producând o deplasare în cazul în care non-respondenții au caracteristici de interes care sunt diferite de cele ale respondenților.
- *Erori de măsurare*, apar atunci când răspunsul oferit diferă de valoarea reală. Astfel de erori pot fi aleatorii, având ca rezultat un grad crescut de dispersie sau abateri sistematice. Ele pot fi atribuite respondentului, anchetatorului, chestionarului, metodei de culegere sau sistemului de păstrare a înregistrărilor aparținând respondentului.
- *Erori de procesare*, apar atunci când se efectuează verificarea, codificarea, introducerea, imputarea și tabularea datelor. La fel ca și erorile de măsurare, erorile de prelucrare pot conduce la dispersie și/sau abateri.

Există două tipuri ale erorilor de procesare. Primul tip se referă la *micro-date*. Al doilea tip privește *macro-datele* și implică *greșeli grave în calculul sau prezentarea agregatelor care nu sunt identificate decât după publicare*. Greșelile sunt erorile cele mai vizibile publicului, primind o atenție negativă sporită. Exemple pot fi atunci când nu se aplică corect metodologia, atunci când se include din neatenție un număr greșit în publicație și când se creează impresii greșite prin prezentări analitice sau diagrame. Acestea pot apărea în oricare etapă de elaborare a statisticii: programare, calculație, scrierea rapoartelor, editarea manuscriselor, etc.

Recomandări

- Determinați gradul de evaluare a calității datelor necesar pentru procesul statistic/ anchetă, luând în considerare următorii factori: utilizarea datelor, riscul de producere a erorilor și impactul erorilor asupra utilizării datelor; variația calității de-a lungul timpului; costul evaluării în raport cu costul total al programului; nevoile de îmbunătățire a calității și/sau eficienței; utilitatea rezultatelor pentru utilizatori și dacă ancheta se repetă sau nu.
- În cazul datelor provenind din recensăminte și anchete, solicitările minime cuprind măsurarea sau evaluarea erorilor de acoperire, a ratei de răspuns sau de imputare și (în cazul unei anchete prin sondaj) măsurarea erorilor de eșantionare pentru indicatorii cheie.
- Realizați evaluarea calitativă pe baza unei opinii avizate, atunci când măsurătorile calitative sunt limitate din cauza naturii rezultatelor, a constrângerilor de timp, cost, precum și a fezabilității tehnice.
- Integrați evaluarea calității datelor în cadrul proiectării generale a anchetei, întrucât informațiile necesare sunt adesea cel mai bine culese în timpul anchetei. Rapoartele privind calitatea datelor vor trebui incluse în programul de diseminare a anchetei.
- În cazul unor anchete repetate nu este necesar și nici posibil să se realizeze o evaluare detaliată a calității cu fiecare ocazie. Cu toate acestea, va trebui programată o revizuire la intervale regulate, să zicem, anual și nu doar atunci când survin anumite probleme.
- Implicați utilizatorii în planificarea evaluării și în dezbaterile rezultatelor. Atunci când împrejurările permit, se recomandă implicarea acestora în procesul efectiv de evaluare.
- Aveți în vedere următoarele metode de evaluare:
 - verificarea coerenței în raport cu sursele externe de informații; de exemplu: alte anchete, repetări ale aceleiași anchete sau date administrative.
 - verificarea coerenței interne, realizată, de exemplu, prin calcularea proporțiilor care sunt cunoscute ca încadrându-se în anumite limite (raportul bărbați-femei, valorile medii ale anumitor proprietăți, etc.)

- analiza celor mai mari unități în mod individual, în ceea ce privește contribuția lor la estimarea generală (aplicată de obicei anchetelor în întreprinderi)
- calcularea indicatorilor de calitate; de exemplu: a ratei de non-răspuns, a ratei de imputare și a coeficienților de variație
- programarea unor ședințe de feedback împreună cu personalul implicat în colectarea și prelucrarea datelor
- verificări realizate de experți bine informați, inclusiv revizuri externe efectuate înaintea publicării.
- Tipul și numărul greșelilor care au fost în mod oficial recunoscute și au avut drept consecință revizuirii, trebuie analizate și prezentate, pentru câțiva ani anteriori.
- Procedurile de minimizare a riscurilor greșelilor și de gestionare a situației când s-au descoperit greșeli, ar trebui dezvoltate și documentate.

Referințe

Biemer, P., Groves, R.M., Mathiowetz, N.A., Lyberg, L. and Sudman, S. (eds.) (1991). *Measurement Errors in Surveys*. Wiley, New York.

Lessler, J.T. and Kalsbeek, W.D. (1992). *Nonsampling Errors in Surveys*. Wiley, New York.

3.13 Controlul confidențialității

Scop și obiective

INS este obligat prin lege să asigure *confidențialitatea* privind datele personale ale respondenților atunci când se publică anumite date. Aceasta înseamnă asigurarea faptului că nici o dată personală a vreunui individ (persoană fizică sau juridică) nu poate fi *dezvăluită*, adică nu poate fi în mod exact dedusă din cadrul unei serii restrânse.

O celulă dintr-un anumit tabel este considerată a fi *sensibilă* dacă aceasta are ca rezultat o *dezvăluire*, adică prezintă date despre o anumită persoană, indiferent de faptul că datele sunt sau nu considerate "sensibile" de către respondent. (De exemplu: în cazul unui respondent înțreprindere, venitul poate fi sensibil, în timp ce punctul de lucru poate să nu fie.)

Controlul confidențialității se referă la măsurile care trebuie luate pentru a modifica datele finale astfel încât confidențialitatea să fie asigurată. Întrucât acesta are de obicei drept rezultat modificarea sau suprimarea unor date, obiectivul este asigurarea confidențialității cu pierderi minime în ceea ce privește rezultatele.

Din perspectiva confidențialității, există o deosebire importantă între:

- Rezultatele micro-datelor – anonimizate pentru înregistrările individuale și
- Rezultatele datelor agregate/tabulare – sub forma tabelelor de frecvență (calculare) sau tabelelor de magnitudine (precum venitul).

Percepția publicului în ceea ce privește vigilența în asigurarea confidențialității este tot atât de importantă ca și ceea ce se realizează de fapt în acest sens.

Principii

INS a elaborat Regulile privind confidențialitatea datelor statistice, o politică pentru protejarea informațiilor clasificate.

Controlul dezvăluirii pentru tabele

Regulile pentru a determina dacă o anumită celulă este sensibilă sau nu au la bază *criteriile de confidențialitate*. Se folosesc două tipuri de criterii. Primul tip depinde de numărul de

respondenți care au contribuit la cifra din celula respectivă iar altul depinde de dimensiunea relativă a contribuțiilor respondenților. Un exemplu pentru primul tip este acela că numărul de respondenți care contribuie la o celulă trebuie să depășească valoarea 3. (Mai puțin de 3 este inacceptabil, de vreme ce fiind doar doi respondenți, atunci unul dintre ei ar putea calcula valoarea celuilalt respondent printr-o simplă scădere.)

Un exemplu general de regulă ce depinde de dimensiunea relativă este *regula (n,k)* în care o celulă este desemnată ca fiind sensibilă dacă cel mai mare număr n de respondenți din ea însumează $k\%$ sau mai mult din valoarea totală a celulei. O altă regulă este aceea a procentului p . De exemplu, conform regulii $p=15$, o celulă este considerată sensibilă dacă suma contribuțiilor celei de-a treia valori și a tuturor valorilor mai mici ale respondenților reprezintă mai puțin de 15% din valoarea cea mai mare a unui respondent. Aceasta asigură că unitatea cu a doua valoare ca mărime nu poate estima prea exact valoarea unității celei mai mari.

Rubricile care sunt considerate sensibile conform criteriilor de confidențialitate vor trebui modificate sau ascunse. Există mai multe reguli posibile de *ascundere a celulelor*. Ascunderea doar a rubricilor sensibile nu este suficientă atunci când totalurile marginale sunt afișate, întrucât valoarea din rubrica ascunsă poate fi dedusă. Pentru a păstra confidențialitatea este necesară executarea *ascunderii complementare*, care determină identificarea tuturor celorlalte celule *complementare* care trebuie ascunse pentru a proteja celula sensibilă. Pentru a identifica celulele complementare, există la dispoziție software specializat.

Există și alte alternative la ascunderea celulelor. Una dintre metode este agregarea categoriilor, astfel încât puține sau niciuna dintre celule să nu fie sensibilă. Alte metode posibile constau în adăugarea unei "perturbări" (noise) în cadrul datelor din tabel sau în micro-datele de bază. Aceasta se poate realiza, de exemplu, prin rotunjire, deși rotunjirea convențională, adică cea efectuată până la cel mai apropiat nivel superior al unității de măsură, poate să nu adauge suficiente "perturbări" pentru a oferi protecție.

Există câteva pachete software pentru a controla confidențialitatea, care include CONFID (Institutul Canadian de Statistică, 2002) și ARGUS (Hundepool et al, 2002).

În cazul unor furnizări multiple din aceeași bază de date, în special în cazul solicitărilor ad-hoc, este dificil să se găsească software-ul care să poată satisface toate cererile de confidențialitate posibile, fiind necesară intervenția manuală.

Controlul dezvăluirii pentru fișiere de micro-date

Chiar și în absența informațiilor de identificare evidente, cum ar fi nume și numere de telefon, fișierele de micro-date conțin un anumit număr de date, denumite *date cheie*, care, luate în combinație pot identifica o anumită persoană. Aceasta reprezintă dezvăluirea efectivă. Chiar dacă indivizii identificați nu sunt chiar unici sau dacă aceștia au fost identificați în mod eronat, *aparența* unei dezvăluiri poate fi dăunătoare pentru INS.

Numărul și natura datelor cheie determină riscul de dezvăluire. Unele caracteristici cum ar fi adresa exactă sau venitul realizat sunt considerate informații cu risc mare de dezvăluire. Fișierele de micro-date ale anumitor întreprinderi sunt rareori făcute publice datorită naturii eterogene ale informațiilor. Prezența unor relații ierarhice între diferitele departamente poate de asemenea crește riscul de dezvăluire. Situația devine și mai complicată în cazul anchetelor longitudinale, întrucât setul variabilelor cheie poate evolua de-a lungul timpului.

Riscul de dezvăluire a informațiilor depinde și de metodologie.

- Acesta crește o dată cu rata de eșantionare. În cazul în care rata de eșantionare este de 100% pentru întreaga populație sau pentru anumite straturi identificabile, micro-datele nu vor trebui făcute publice.
- Prezența erorilor de măsurare sau a valorilor imputate scade riscul de dezvăluire.

Există două *metode principale de restricționare a datelor* în vederea controlului confidențialității micro-datelor.

- *Metodele de reducere a datelor* care cuprind eșantionare, lărgirea categoriilor de clasificare, eliminarea datelor selectate pentru unii respondenți și eliminarea unor respondenți selectați.
- *Metode de modificare a datelor* includ adăugarea unei "perturbații" aleator; schimbarea datelor între înregistrări, înlocuirea unor grupuri mici de unități cu valori medii și înlocuirea unor date pentru anumiți respondenți cu valorile imputate.

Ca o alternativă la combaterea scurgerii de micro-date, pot fi inițiate proceduri de *restricționare a accesului*, care permit accesul la aceste micro-date doar cercetătorilor, în condiții controlate.

Recomandări

Datele tabulare

- Determinați sensibilitatea fiecărei celule. Nu faceți public un tabel de date dacă acesta conține valori care sunt considerate a fi sensibile.
- Determinați dacă celulele de frecvență zero reprezintă o problemă. Celulele de frecvență zero pot scoate la iveală informații sensibile din tabelele de date de magnitudine.
- Aveți în vedere metode alternative pentru ascunderea celulelor, de exemplu, utilizarea metodei de rotunjire aleatoare.
- În cazul unor furnizări multiple dintr-o singură bază de date, folosiți-vă experiența pentru a suplimenta rezultatele obținute cu ajutorul programului de protecție a confidențialității.

Micro-datele

- Evaluați riscul de dezvăluire a informațiilor din fișierele cu micro-date și aplicați metode corespunzătoare de reducere a datelor sau de modificare a lor.
- Asigurați-vă de faptul că toate fișierele cu micro-date care vor fi făcute publice sunt verificate de o comisie de experți din cadrul INS înainte de a fi publicate.
- În cazul anchetelor longitudinale, se va defini strategia de publicare a datelor și cea de confidențialitate, înainte de primul val. Aceasta poate implica adoptarea unor ipoteze privind datele cheie și cum vor evolua acestea de-a lungul timpului.
- Ca alternativă la punerea la dispoziție a micro-datelor, aveți în vedere folosirea unor metode de restricționare a accesului, cum ar fi accesarea de la distanță sau centrele de cercetare a datelor.
- Folosiți un software special pentru controlul confidențialității informațiilor în loc de sistemele de protecție obișnuite, ori de câte ori este posibil acest lucru.

Regulile de confidențialitate ale INS

Pentru a preveni diseminarea datelor statistice confidențiale, trebuie respectate următoarele reguli:

- Datele statistice individuale privitoare la o persoană fizică sau juridică nu trebuie diseminate dacă persoana în cauză nu își exprimă acordul în scris.
- Datele agregate privind două persoane fizice sau juridice nu pot fi diseminate dacă acestea nu își exprimă acordul în scris.
- Datele statistice agregate pentru trei sau mai multe persoane fizice sau juridice pot fi diseminate, dar numai cu condiția ca aceste date să nu permită mai departe identificarea datelor individuale. În special, aceste date nu pot fi diseminate dacă o persoană juridică sau fizică reprezintă peste 80% din valoarea sau cantitatea totală a unităților din celulă.
- Identificatori unici: codul personal, codul fiscal, denumirea și adresa, localitatea și alte informații care nu pot fi depersonalizate (nu pot fi anonimizate) nu vor fi incluse în fișierele de date individuale pentru utilizarea publică.

Accesul la datele statistice individuale poate fi permis, în scop științific, cu condiția să se fi luat toate măsurile de protejare privind păstrarea confidențialității datelor statistice și supus:

- Aprobării conducerii INS și Comisiei pentru confidențialitatea statistică, stabilită prin Ordinul Președintelui INS;
- Aprobării conducerii Direcției regionale/județene de statistică și aprobării Comisiei pentru confidențialitatea statistică, stabilită la nivel teritorial prin decizia directorului executiv al Direcției regionale/județene de statistică.

Referințe

Eurostat (1999) *Manual on Disclosure Control Methods*

Brackstone, G. and White, P. (2002). *Data stewardship at Statistics Canada*. Proceedings of the Social Statistics Section, American Statistical Association, 284-293.

Eurostat (1996). *Manual on Disclosure Control Methods*. Luxembourg: Office for Official Publications of the European Communities.

Hundepool, A., van de Wetering, A., de Wolf, P.-P., Giessing, S., Fischetti, M., Salazar, J.-J. and Caprara, A. (2002). *ARGUS user manual 2.1*. Statistics Netherlands, Voorburg.

Institutul Național de Statistică, *România Reguli privind Confidențialitatea Datelor Statistice*

Institutul Național de Statistică, *România Politica privind Protecția Informațiilor Clasificate în Cadrul Institutului Național de Statistică*

3.14 Analiza și prezentarea datelor

Scop și obiective

Analiza datelor este procesul de transformare a datelor în informații. Rezultatele analitice subliniază utilitatea surselor de date și scot în evidență anumite aspecte. Rezultatele analitice pot fi de și mai mare importanță atunci când, din motive de confidențialitate, nu este posibilă transmiterea micro-datelor. Acestea sunt în mod specific prezentate sub formă de comentarii atașate rezultatelor și de articole publicate separat.

Principalele etape din cadrul procesului analitic constau din identificarea problemelor, punerea unor întrebări semnificative, oferirea unor răspunsuri prin examinarea și interpretarea datelor și comunicarea mesajelor către utilizatori.

Analizele joacă un rol important în dezvoltarea, revizuirea și îmbunătățirea anchetei, contribuind la identificarea problemelor și a potențialelor îmbunătățiri. Aceasta facilitează extragerea eficace de informații din anchetele anterioare și din studiile pilot, planificarea de noi activități statistice, furnizarea de informații privind lipsa unor date, pentru a realiza anchete și pentru a formula obiectivele calității.

Prezentarea datelor și a comentariilor ce însoțesc articolele analitice asociate necesită luarea în calcul a celor mai bune căi de comunicare cu utilizatorii.

Principii

Rezultatele analizei pot fi clasificate în două mari tipuri generale:

- *rezultate descriptive*, legate de populația anchetei la momentul când au fost colectate datele,
- *studiile* care depășesc nivelul segmentului de populație anchetat.

Comentariile atașate rezultatelor datelor se reduc la rezultatele descriptive. Mandatul INS nu sprijină în prezent studiile analitice care interpretează rezultatele. Aceste studii pot fi conduse de cercetători independenți, sub îndrumarea INS.

Orice concluzie prezentată în comentariu sau în studiul analitic, în special acelea care au un impact asupra politicii publice, va trebui sprijinită de datele care se află în curs de analizare.

Standardele de prezentare trebuie dezvoltate și respectate.

Recomandări

Pregătirea articolelor analitice

- Identificați aspectele de interes (atât pe acelea curente cât și pe acelea care este posibil să apară în viitor), publicul țintă, sursa de informație și metodele de analiză.
- Asigurați-vă de faptul că informațiile corespund analizei, de exemplu, de faptul că populația cadru oferă o aproximare suficient de apropiată de populația țintă, de faptul că definițiile și conceptele sunt relevante, că natura longitudinală sau transversală a anchetei este corespunzătoare, că precizia datelor este suficientă pentru a se obține rezultate concludente.
- În cazul în care este folosită mai mult de o sursă de date pentru analiză, faceți o cercetare privind consecvența informațiilor provenite din acea sursă și analizați modul în care acestea pot fi combinate în mod corespunzător.
- Aveți în vedere modul în care vor trebui gestionate non-răspunsurile pe variabile și dacă vor trebui incluse valori imputate.
- Atunci când faceți o comparație între două grupe de persoane sau de întreprinderi, sau alte unități, controlați factorii exteriori. Dacă se constată diferențe semnificative între grupuri, atunci luați în considerare explicații alternative plauzibile pentru aceste diferențe.
- De vreme ce analizele se bazează mai degrabă pe observații decât pe rezultatele unui experiment controlat, evitați să trageți concluzii privind cauzalitatea.
- Nu vă concentrați atenția asupra trendurilor pe termen scurt fără a le inspecta în lumina celor pe termen lung și mediu. În mod frecvent, trendurile pe termen scurt sunt numai fluctuații minore în jurul mai importantelor trenduri pe termen mai lung.
- Atunci când este posibil, evitați reperele de timp arbitrare, cum ar fi schimbările de la an la an. În schimb, folosiți repere semnificative, cum ar fi ultimul punct de cotitură în ceea ce privește datele din economie, diferențele dintre generații privind statistica demografică și modificările legislative pentru statistica socială.

Prezentarea articolelor analitice

- Înainte de a scrie un articol, redactați o schiță a articolului pentru a indica aspectele care trebuie abordate, datele utilizate, metodele analitice, rezultatele care vor trebui scoase în evidență și cele mai interesante constatări.
- Axați articolul pe indicatorii și subiectele importante. Încercarea de a fi prea cuprinzător tinde să interfereze cu un material cursiv.
- Aranjați ideile într-o ordine logică și în ordinea importanței sau relevanței lor. Folosiți titluri, sub-titluri și note explicative pentru a consolida structura articolului.
- Folosiți un limbaj cât mai simplu, atât cât permite subiectul. În funcție de publicul-țintă al articolului, pierderea preciziei poate fi o concesie pentru un text mai ușor de citit.
- Utilizați grafice în cursul textului și tabele pentru a comunica mesajul. Folosiți titluri care reușesc să surprindă esența articolului, ca de exemplu "Câștigurile femeilor sunt mai reduse decât cele ale bărbaților", în locul unor titluri tradiționale, cum ar fi "Câștigurile în funcție de vârstă și sex".

- Urmăriți să ajutați cititorii să înțeleagă informațiile din tabele și grafice prin dezbaterile lor în cadrul textului.
- Atunci când utilizați tabele, țineți cont de faptul că formatul general contribuie la claritatea datelor și reduce posibilitatea de interpretare greșită. Acest lucru cuprinde spațiere adecvată, redactarea, plasarea și modul de prezentare a titlurilor, a rândurilor, capetelor de tabel și a notelor de subsol.
- Explicați practicile și procedurile de rotunjire. Pentru prezentarea unor date rotunjite, nu folosiți mai multe cifre decât este necesar pentru acuratețea informației.
- Atunci când prezentați detalii despre anumite rate, aveți grijă să faceți diferența dintre schimbările procentuale și schimbările dintre punctele procentuale. Definiți baza utilizată pentru rate.
- Asigurați-vă de faptul că bibliografia este prezentată în mod exact și că se fac trimiteri la aceasta în cadrul textului.
- Verificați dacă există greșeli, asigurați-vă de acuratețea cifrelor folosite în text, tabele și grafice și de exactitatea datelor externe.
- Asigurați-vă că obiectivele pe care le-ați menționat în introducere sunt îndeplinite iar concluziile corespund cu realitatea.
- Oferiți articolul spre a fi revizuit de către cel puțin două persoane.
- Pregătiți o prezentare a articolului în fața colegilor înainte de a face expunerea în fața unei asistențe externe.

Comentarii

Recomandările de mai sus se pot aplica și comentariilor legate de rezultatele datelor, cu condiția ca analiza să fie mai puțin profundă și să se concentreze mai degrabă asupra rezumării rezultatelor principale, decât asupra problemelor de analiză.

Referințe

Chambers, R.L. and Skinner, C.J. (eds.) (2003). *Analysis of Survey Data*. Wiley, Chichester.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Duxbury Press.

3.15 Diseminarea datelor

Scop și obiective

Diseminarea reprezintă furnizarea de date statistice utilizatorilor. Mecanismele de difuzare on-line includ Internetul, bazele de date statistice și fișierele web, interviurile prin televiziune și radio și conferințele de presă. Mecanismele offline includ publicațiile tipărite, CD-ROM-urile, comunicatele de presă, faxul, poșta electronică, răspunsurile la cerințele ad-hoc și prezentările la evenimente.

Principii

Datele diseminate vor trebui să fie exacte, complete, accesibile și tarifate adecvat, ușor de înțeles, utilizabile, oferite în timp real și care să respecte cerințele de confidențialitate. Managerii responsabili cu diseminarea vor trebui să aibă în vedere necesitățile și așteptările utilizatorilor bazate pe feedback-ul utilizatorilor, pe testarea produselor și pe activitățile de marketing. Aceștia vor trebui să exploateze noile tehnologii pentru a permite utilizatorilor să aibă un acces optim la informație și să o proceseze eficient și cu ușurință.

Datele și perioadele în care se vor difuza informațiile trebuie anunțate dinainte. Utilizatorilor li se va permite în mod egal accesarea datelor. Datele trebuie publicate pe web, transmise

electronic, sau difuzate pe suport hârtie, exact în același timp și în conformitate cu anunțurile prealabile.

Rezultatele trebuie însoțite de *metadate* care permit utilizatorilor să determine cât de potrivite sunt datele pentru anumitele utilizări pe care aceștia le doresc. Acestea cresc de asemenea încrederea pe care o au utilizatorii în date. Metadatele furnizate trebuie să cuprindă concepte și definiții, populații țintă și eșantioane de populație, proiectarea eșantioanelor, chestionare, metode de imputare, de culegere și estimare a datelor, ratele de răspuns, posibile erori și împrejurări excepționale. Nivelul de detaliu trebuie să fie adecvat cerințelor utilizatorilor. Acestea pot determina elaborarea de rezumate a metadatelor vizibile inițial prin trimiteri la metadate mai detaliate pentru utilizatorii care au nevoie de ele.

Recomandări

- Asigurați-vă de faptul că datele sunt verificate și validate de către o altă unitate decât cea care le produce. Unitatea care verifică trebuie să aibă experiență dovedită.
- Dezvoltați și utilizați un format comun pentru toate produsele din aceeași categorie sau temă și dezvoltați un standard al agenției.
- Asigurați-vă că datele difuzate sunt exact aceleași ca și cele prezente în bazele de date interne. Trebuie să fie posibilă întotdeauna reproducerea rezultatelor întocmai.
- Testați orice produs electronic înainte de difuzare pentru a vă asigura că acesta funcționează conform așteptărilor utilizatorilor.
- Furnizați metadate care să conțină descrieri corespunzătoare ale calității și metodologiei.
- Furnizați indicatori de calitate ai datelor sau instrumente care să sprijine utilizatorii în evaluarea datelor, de exemplu tabele de vizualizare a CV-urilor, programe de calcul a gradului de dispersie a eșantionului.
- Definiți cu claritate regulile, condițiile și termenii drepturilor de autor pentru utilizarea sau re-diseminarea informației statistice.
- Însoțiți fiecare rezultat cu date de contact, număr de telefon și o adresă de e-mail.
- Utilizați la maximum date generalizate și sisteme de diseminare a metadatelor.
- Asigurați-vă că personalul implicat în procesele de diseminare a datelor este bine instruit.
- Analizați și utilizați standardele și cele mai bune practici de diseminare din țările UE.

Referințe

Institutul Național de Statistică, România (2006), *Politica privind Diseminarea Datelor*.

3.16 Documentarea

Scop și obiective

Documentarea se referă la descrierea unei anchete/proces statistic, cuprinzând conceptele, definițiile și metodele utilizate pentru elaborarea datelor, a rezultatelor și factorilor care influențează calitatea acestora. Documentația reprezintă o modalitate de comunicare în timpul elaborării și implementării unei anchete, cuprinzând nu doar deciziile care au fost luate, dar și motivul pentru care au fost luate. Documentarea sprijină perfecționarea ulterioară a anchetei în cazul celei periodice și dezvoltarea anchetelor similare. De asemenea, documentația stă la baza informării utilizatorilor în legătură cu metodologia și calitatea unei anchete. Documentarea poate fi realizată pentru mai multe grupuri diferite, incluzând managerii, personalul tehnic, cei însărcinați cu planificarea altor anchete și utilizatorii.

Principii

Documentația urmărește să ofere o înregistrare completă, clară și multi-funcțională a anchetei și a rezultatelor, actualizate, bine organizate, ușor de găsit, concise și precise. Nivelul de detaliere depinde de publicul căreia îi este destinată, de sursele de date, de proiectarea eșantionului, de metodele de culegere și diseminare, de utilizatorii și utilizările datelor și de bugetul total.

Accesul efectiv și prezentarea sunt importante. Documentarea poate fi multi-media, de exemplu pe suport de hârtie și electronic. Diferitele documente pot fi pregătite pentru un public divers și pentru scopuri diferite.

Există numeroase standarde privind conținutul, dintre care cele mai importante în contextul INS sunt *Regulile privind elaborarea și editarea metodologiei anchetei*, *Regulile privind elaborarea și editarea lucrărilor statistice* ale INS, elaborate de Unitatea de calitate și *Structura Euro SDMX a metadatelor (ESMS)*.

ESMS conține o descriere și o prezentare a metadatelor pentru a fi utilizate în documentarea datelor statistice și în evaluarea proceselor de producere a datelor și a rezultatelor. Conceptele ESMS utilizate sunt compatibile cu acelea ale standardului *Schimbului de date statistice și metadata (SDMX)* și cu terminologia comună din *Vocabularul comun de metadata SDMX*. ESMS va fi implementat la Eurostat și la nivel național.

Recomandări

- Pregătiți documentația anchetei la un nivel de detaliere adecvat publicului țintă.
- Indicați orice schimbări legate de obiectivele, oportunitatea, frecvența și obiectivele calitative ale datelor care au avut loc pe parcursul elaborării anchetei, de exemplu, datorită constrângerilor bugetare, fezabilității percepute, rezultatelor noilor studii pilot sau noilor tehnologii.
- Indicați utilizarea standardelor internaționale și naționale și evidențiați abaterile.
- Indicați rolul comisiei de experți și al utilizatorilor în elaborarea și desfășurarea anchetei.
- Pregătiți versiuni ale documentației metodologice, accentuând aspecte diferite pentru diferiții cititori. Furnizați directorilor generali și utilizatorilor o privire metodologică de ansamblu. Oferiți experților o relatare cuprinzătoare și consolidată a procedurilor și a problemelor tehnice.
- Furnizați documente care să însoțească produsele și supuneți-le unor analize extinse din partea colegilor și a managerilor pentru a vă asigura de relevanța, acuratețea și lizibilitatea lor.
- Respectați politicile de revizuire a documentației.

Referințe

Regulile privind elaborarea și editarea metodologiilor cercetărilor statistice

Regulile privind elaborarea și editarea lucrărilor statistice

Structura de Metadata Euro SDMX (ESMS).

Sistemul de Schimb de Date și Metadata Statistice (SDMX).

Vocabularul Comun de Metadata.

3.17 Utilizarea datelor administrative

Scop și obiective

Datele administrative sunt datele colectate în timpul administrării programelor guvernamentale. Procedurile administrative includ colectarea impozitelor din venituri și TVA, reglementarea circulației bunurilor și a persoanelor dincolo de granițe, înregistrarea nașterilor, a căsătoriilor și a deceselor, administrarea pensiilor, oferirea serviciilor medicale. Datele administrative sunt colectate în conformitate cu legislația în vigoare. Există două tipuri principale:

- Date de înregistrare, care conțin detaliile tuturor unităților supuse procedurii administrative respective;
- Date de tranzacționare, care conțin detaliile interacțiunii acelor unități cu procedura administrativă.

Fiecare dată înregistrată se referă la un individ –persoană, contribuabil, pensionar etc.– iar identitatea unei unități care corespunde unui anumit registru este extrem de importantă. Dimpotrivă, în cazul *datelor statistice*, identitatea indivizilor nu mai are o valoare intrinsecă o dată ce datele au fost colectate.

Unele date administrative sunt înregistrate în formă tabelară și diseminate direct de INS. De exemplu, datele legate de bunurile expediate sunt obținute de la vamă, prelucrate și diseminate ca fiind date de comerț exterior.

Pe lângă tabelarea și diseminarea directă, datele administrative mai pot fi folosite:

- Pentru a elabora și a întreține registrul întreprinderilor;
- Pentru a înlocui, parțial sau în totalitate, colectarea de date, de exemplu utilizarea datelor de impozitare pentru întreprinderile mici în locul colectării datelor legate de venituri și cheltuieli;
- Pentru a ajuta în procesul de editare și imputare;
- Ca informații auxiliare în estimări – în scopuri de calibrare și/sau întocmirea unor calendare;
- Pentru evaluarea anchetelor, mai precis *confruntarea datelor* și compararea datelor administrative și a celor din cercetări statistice.

Principii

Avantaje

Avantajele utilizării datelor administrative sunt, în primul rând, furnizarea fără costuri de către agențiile administrative responsabile pentru culegerea lor, conform legii. În al doilea rând, ele nu impun o sarcină statistică suplimentară de răspuns asupra persoanelor, la care se referă datele. În al treilea rând, tehnologia modernă permite INS să acceseze cu rapiditate și să prelucreze seturi mari de date administrative.

Dezavantajele datelor administrative constau în faptul că acestea sunt culese mai degrabă în scopuri administrative, decât în scopuri statistice. De obicei, potențialii utilizatori statistici de date administrative nu sunt cunoscuți sau sunt considerați neimportanți la momentul în care programul administrativ este organizat. Aceasta înseamnă că INS are un control limitat sau nu are control deloc asupra gradului de acoperire, conținutului, preciziei sau sincronizării. Modificările ulterioare în legislație pot determina modificări ale datelor, în afara controlului INS.

Pe scurt, într-un moment în care cererea de date este în creștere, există constrângeri bugetare serioase, iar preocupările privind ratele de răspuns și sarcina de răspuns sunt în creștere, utilizarea datelor administrative, reprezintă o alternativă atractivă și/sau complementară la culegerea directă de date. Astfel politica INS este aceea de a utiliza datele administrative atunci când acestea prezintă o alternativă ce implică costuri mici, sau un supliment la culegerea directă de date.

Ca și în cazul oricărui alt program de achiziție de date, deciziile legate de folosirea unei surse de date administrative trebuie să urmeze unui proces de evaluare a datelor, în ceea ce privește gradul de acoperire, conținutul, conceptele și definițiile, frecvența, punctualitatea, procedurile de asigurare a calității, precum și stabilitatea programului administrativ în timp și costurile pe care le implică obținerea și prelucrarea datelor.

În prelucrarea datelor administrative INS trebuie să realizeze efectiv toate sub-procesele, privind datele statistice discutate în secțiunile anterioare. Astfel, recomandările din secțiunile anterioare sunt în mare parte aplicabile în mod egal înregistrărilor administrative. Chiar și recomandările privind eșantionarea și captarea de date pot fi relevante dacă datele administrative există doar pe hârtie și trebuie să fie eșantionate, codificate și captate. Totuși, există câteva considerații suplimentare, care reprezintă subiectul acestei secțiuni.

În primul rând, folosirea datelor administrative poate da naștere unor preocupări legate de confidențialitate. Aceste preocupări au o importanță cu atât mai mare, atunci când datele administrative sunt legate de alte surse de date. Astfel, asemenea legături între diferite registre sunt, de obicei, controlate cu grijă.

În al doilea rând, combinarea datelor care provin dintr-o sursă administrativă cu cele care provin din altă sursă nu este ușoară în absența unui sistem de identificare comun care să acopere ambele surse. Sunt necesare tehnici de *interconectare a registrelor*. Dacă scopul conectării surselor este crearea unui cadru și întreținerea, editarea și imputarea sau ponderarea, îmbinarea exactă este esențială. Dacă, pe de altă parte, sursele sunt conectate în scopul analizei datelor, *combinarea statistică*, adică îmbinarea registrelor cu proprietăți statistice asemănătoare poate fi suficientă.

Recomandări

- Creați și mențineți legături strânse cu persoanele care gestionează sursele administrative. Legătura este necesară nu numai în negocierea utilizării datelor administrative, ci și pentru ca modificările procedurilor administrative să nu ia prin surprindere INS, acesta putând chiar să influențeze aceste modificări. Feed-back-ul, către gestionarii datelor, în privința problemelor și sugestiilor poate întări sursele administrative. Ocazional, biroul de statistică poate ajuta la includerea cerințelor statistice în noile programe, înainte de demararea acestora. Astfel de oportunități sunt rare, dar atunci când se ivesc, valoarea potențială pentru INS prevalează asupra timpului și resurselor necesare pentru colaborare.
- Înțelegeți contextul în care organizația administrativă a creat programul administrativ (ex.: legislația, obiectivele și necesitățile). Acesta are un impact deosebit asupra acoperirii, conținutului conceptelor și definițiilor, frecvenței și punctualității, calității și stabilității în timp.
- Studiați fiecare dată din sursa administrativă care va fi folosită în scopuri statistice. Cercetați-i calitatea. Înțelegeți conceptele, definițiile și procedurile care stau la baza colectării și procesării ei. Unele variabile pot avea o calitate slabă și nu vor putea fi utilizate. De exemplu, calitatea codurilor de clasificare (ocupația, activitatea industrială, geografia) poate fi insuficientă pentru scopuri statistice sau poate restrânge posibilitatea de folosire. În asemenea cazuri, identificați modalități de a pune la dispoziție personalul INS sau resurse pentru îmbunătățirea calității.
- Alegeți metodologia de anchetă care se potrivește unei anumite surse administrative.
- La fel ca la datele colectate în cadrul unei anchete și la datele administrative poate exista non-răspunsul sau răspunsul parțial. În unele cazuri, lipsa unei coordonări în obținerea datelor în mod eficient are ca rezultat non-răspunsul. Trebuie dezvoltate astfel procedurile de imputare și ponderare.
- Imputarea și/sau transformarea pot fi necesare și în cazurile în care unele dintre unități raportează datele cu o frecvență diferită față de cea necesară pentru anchetă.

- Stările activității unităților conform surselor administrative sunt deseori diferite față de cele cerute pentru scopurile statistice. De pildă, agenția de impozite va reține în înregistrările sale o întreprindere falimentară cât timp aceasta datorează impozit, în timp ce, pentru biroul de statistică întreprinderea este inactivă. Trebuie acordată astfel o atenție specială identificării unităților administrative, care sunt aparent active, dar din perspectiva statistică sunt inactive. De exemplu neplata taxei pe valoarea adăugată timp de 18 luni poate fi considerată o dovadă a inactivității.
- Luați în considerare faptul că pot exista erori de măsurare semnificative dacă datele administrative au implicații financiare. De exemplu, venitul poate fi subevaluat și/sau cheltuielile sunt supraevaluate pentru a reduce obligațiile fiscale.
- Aveți în vedere implicațiile de confidențialitate în cazul publicării informațiilor din registrele administrative. Deși INS are dreptul să acceseze registre administrative în scopuri statistice, este posibil ca această utilizare să nu fi fost prevăzută de furnizorii inițiali ai informațiilor. Din acest motiv, managerii anchetelor ar trebui să fie pregătiți să explice și să justifice valoarea publică și natura nevătămătoare a acestei utilizări în scopuri statistice.
- Efectuați o evaluare continuă sau periodică a calității datelor intrate. Este important să vă asigurați că se menține calitatea datelor, deoarece INS nu controlează procesul de colectare a datelor. Evaluarea poate conține metode de control al calității statistice, comparații cu alte surse și studii de urmărire a eșantioanelor.
- Preocupările legate de confidențialitate, care pot apărea atunci când este utilizată o singură sursă administrativă se multiplică atunci când se face cuplarea cu alte surse. Este necesară o politică de relaționare a registrelor pentru a se asigura că valoarea cuplajului prevalează asupra oricărei nerespectări a confidențialității pe care ar putea-o determina.
- Atunci când datele din două sau mai multe surse sunt comparate fără o schemă comună de identificare, aveți în vedere metodele de combinare statistică, dacă nu este necesară o îmbinare exactă.

Referințe

Cox, L.H. and Boruch, R.F. (1988). *Record Linkage, Privacy and Statistical Policy*. Journal of Official Statistics, 4, 3-16.

Eurostat (2001) *Use of Administrative Sources for Business Statistics Purposes*

Hidiroglou, M.A., Latouche, M.J., Armstrong, B., and Gossen, M. (1995). *Improving Survey Information Using Administrative Records: The Case of the Canadian Employment Survey*. Proceedings of the Annual Research Conference, U.S. Bureau of the Census, 171-197.

Michaud, S., Dolson, D., Adams, D., and Renaud, M. (1995). *Combining Administrative and Survey Data to Reduce Respondent Burden in Longitudinal Surveys*. Proceedings of the Section on Survey Research Methods, American Statistical Association, 11-20.

4 Cadrul General de Management al Calității

Remarci introductive

Acest capitol schițează *Cadrul General de Management al Calității* (QMF) în cadrul căruia se aplică principiile calității descrise în Capitolul 3. În timp ce Capitolul 3 se referă la anchete și este destinat managerilor de anchetă, Capitolul 4 se referă la instituție ca întreg și este destinat managerilor seniori și personalului cu responsabilitatea specială de a promova managementul calității în cadrul organizației.

Managementul calității implică identificarea și urmărirea obiectivelor calității în lumina cererii utilizatorilor pentru mai multă informație și a altor considerații privind calitatea, luând în considerare resursele financiare și umane și sarcina de răspuns. Astfel, QMF cuprinde mai multe mecanisme la niveluri diferite în cadrul INS. Eficacitatea acestuia depinde de profesionalismul și motivația personalului și de sinergia ce rezultă din operațiunile celor implicați, care acționează colectiv, având un set comun de obiective privind calitatea.

QMF este prezentat în contextul *componentelor calității rezultatelor*, definite în Capitolul 2, cu secțiuni referitoare la *Managementul relevanței, acurateței, punctualității, accesibilității, comparabilității și coerenței*, împreună cu secțiunile transversale privind *Raportarea calității și Managementul mediului instituțional*.

4.1 Managementul Relevanței

Relevanța reprezintă gradul în care rezultatele statistice corespund nevoilor actuale sau potențiale ale utilizatorilor. Ea depinde de furnizarea datelor statistice care sunt necesare și de măsura în care conceptele utilizate (definiții, clasificări etc.) reflectă nevoile utilizatorilor.

Programul și rezultatele INS trebuie să reflecte în mod corespunzător și continuu nevoile cele mai importante privind datele, pe măsură ce acestea evoluează de-a lungul timpului. Managementul relevanței are astfel două aspecte – determinarea nevoilor statistice și adaptarea programului INS pentru a atinge aceste nevoi.

Determinarea nevoilor utilizatorilor

Cerințele SSE sunt specificate în regulamentele UE. Ele sunt obligatorii. Mai mult, rapoartele de calitate cerute de reglementări presupun în mod specific dovezi că datele cerute sunt produse. Alte nevoi pot veni din partea agențiilor internaționale, în special de la IMF, OECD, UNDP, FAO, etc.

Nevoile naționale sunt identificate prin legătura cu principalii utilizatori, în cea mai mare parte, departamente guvernamentale. Există de asemenea, un feedback din partea altor utilizatori în privința rezultatelor existente culese de Biroul de diseminare. Cercetătorii și analiștii pot indica lipsuri și limitări ale datelor prin intermediul articolelor publicate. Comentariile și criticile din partea media pot fi de asemenea informative.

Trebuie ținut cont de asemenea, de nevoile programelor interne, precum Sistemul Conturilor Naționale, care extrag date dintr-o serie de anchete, le integrează și le reconciliază.

Procesele care determină nevoile de date au ca rezultat și monitorizarea altor componente ale calității și identificarea cerințelor corespunzătoare ale utilizatorilor.

Adaptarea programului

Adaptările programului se fac anual, prin definirea și implementarea Programului Statistic Național Anual (PSNA). Programul de activitate definește rezultatele, sarcinile și oportunitatea pentru activitățile statistice din cadrul INS finanțate de guvern. În fiecare an Programul de activitate pentru următorul an este realizat pe baza celui pentru anul curent, dar adaptat la

nevoile în continuă schimbare. Programul de activitate se transmite apoi la utilizatorii primari – departamente și agenții de stat – pentru comentarii. Pe baza feedback-ului primit și în lumina priorităților de politică curentă, managementul superior al INS finalizează Programul de activitate.

În practică, există constrângeri puternice asupra ajustărilor programului. În primul rând, o mare parte a resurselor bugetare ale INS este alocată anchetelor continue și a proceselor statistice care sunt *nediscreționare* în sensul că ele sunt dictate de reglementările UE sau furnizează informații de bază privind societatea și economia românească, astfel încât nu pot fi schimbate în mod semnificativ. În al doilea rând, este extrem de dificil să se întrerupă, sau chiar să restrângă o anchetă existentă din cauza faptului că utilizatorii vor avea obiecții. În al treilea rând, în esență nu există perspective de a-și asigura finanțările suplimentare de la stat pentru noile programe, decât dacă aceste finanțări sunt de fapt incluse în legislație.

Astfel, reducerile programului sunt minimale, iar extinderile pentru a răspunde nevoilor de noi date sunt în general foarte limitate și inițiate doar prin devierea resurselor alocate altor activități curente.

În fiecare an unele activități statistice sunt finanțate din alte surse, de pildă de către UE, Banca Mondială și/sau alți parteneri internaționali și naționali. Totuși, acestea sunt în mod specific destinate îmbunătățirii proceselor statistice.

Considerente de îmbunătățire

Procedurile actuale pentru gestionarea relevanței pot fi îmbunătățite în trei moduri. Primul, introducerea unui proces de planificare multi-anuală ar conferi mai multe posibilități pentru antrenarea modificărilor în program.

Cel de-al doilea, revizuirile regulate auto-administrate ale tuturor anchetelor ar asigura un feedback mai puternic privind sensibilitatea publicului și relevanța la nivelul anchetei și oportunitățile pentru schimbări sau reduceri ale modulelor de întrebări sau dimensiunilor eșantionului. Aceste revizui sunt prevăzute începând cu 2010, așa cum vor fi descrise în secțiunea privind raportarea calității.

Al treilea mod, opțiunile privind alocarea de resurse pentru inițierea de noi anchete sau extinderea celor existente, ar putea fi explorate. Acestea includ obținerea unei mai bune înțelegeri privind costurile la nivel de anchetă individuală, astfel încât oportunitățile de eficientizare să fie mai ușor identificate, precum și măsurarea sarcinii de răspuns, ca punct de pornire în reducerea dimensiunii chestionarului și a eșantionului.

4.2 Managementul Acurateței

Acuratețea rezultatelor statistice în accepțiune statistică generală reprezintă gradul de apropiere a estimărilor de valorile reale.

Aspecte privind proiectarea

Acuratețea, punctualitatea și coerența obținute în rezultatele anchetei depind foarte mult de proiectarea metodologiei, de procesele de asigurare a calității folosite pentru identificarea și controlul potențialelor erori. Există două aspecte: capacitatea metodologică și politicile și procedurile utilizate de obicei și aplicarea acestei capacități, politici și proceduri în anchetele individuale.

În prezent nu există o unitate de metodologie sau un centru primar de expertiză. Responsabilitatea asupra metodelor de eșantionare și estimare revine celor trei unități responsabile de statistica întreprinderilor, statistica gospodăriilor și de statisticile agricole. Toate celelalte aspecte metodologice, de exemplu proiectarea chestionarului, procedurile de editare, procedurile de imputare, detectarea valorilor aberante sunt gestionate de către statisticieni

responsabili cu anchetele individuale. Direcția de calitate are responsabilitatea de a promova practicile privind calitatea, de pildă raportarea calității și producerea și utilizarea acestor principii ale calității.

În ceea ce privește procedurile metodologice, acest ghid descrie practicile specifice, metodele și considerațiile, care trebuie luate în calcul atunci când se elaborează programele, și indică unde există standarde oficiale. În afară de acest document există câteva proceduri metodologice generale ale INS, de pildă proiectarea și formatarea standardizată a chestionarului. Există practica standard de a pretesta și de a face teste pilot pentru toate chestionarele noi înainte de utilizarea lor în producție. Personalul se bazează pe literatura SSE și pe cea internațională, precum și pe experți pentru îndrumare, de exemplu Ghidul SSE privind ajustarea sezonieră.

În ceea ce privește aplicarea metodologiei, deciziile sunt luate în mare parte de către conducătorii de anchete individuale și de către personalul acestora. Managerii, pe lângă deciziile privind obiectivele de acuratețe acceptate pe baza cunoașterii cerințelor utilizatorului, bugetului și operațiilor, determină și metodele specifice ce vor fi aplicate.

Aspecte de implementare

Acuratețea cercetării statistice depinde de planificare, de procedurile operaționale, de supervizare, de relațiile cu respondenții, de interviuri și instruire. Calitatea și mecanismele de monitorizare a performanței trebuie încadrate în proiectarea anchetei. Sunt necesare două tipuri de informații. Prima informație este necesară monitorizării și corecției imediate a oricăror probleme operaționale. Cea de-a doua informație este necesară pentru evaluarea gradului în care ancheta a fost implementată conform planurilor, sau a gradului în care unele aspecte ale proiectării s-au dovedit problematice și a lecțiilor învățate din punctul de vedere operațional pentru a sprijini proiectele viitoare.

Aspecte de evaluare

Evaluarea acurateței trebuie luată în considerare în etapa de proiectare, întrucât trebuie obținute măsurători pe parcursul desfășurării anchetei. Gradul și complexitatea măsurătorilor depind de dimensiunea anchetei, de buget și de importanța lor pentru utilizatori. Informarea utilizatorilor privind calitatea datelor necesită evaluarea a cel puțin următoarele aspecte: gradul de acoperire al anchetei, eroarea de eșantionare, ratele de non-răspuns și descrierile oricăror probleme sau consecințe grave.

Considerente privind îmbunătățirea

Posibilitățile de îmbunătățire ulterioară în managementul acurateței includ întărirea expertizei în eșantionare și estimare, în proiectarea chestionarului, editarea, imputarea și analiza, includerea calității și mecanismele de monitorizare a performanței în procesele anchetei, introducerea programului de rapoarte de calitate orientate către utilizator (discutate în următoarea secțiune), încurajarea utilizării listei de verificare a anchetei (DESAP) și utilizarea noilor tehnologii și a inovațiilor (cu testarea adecvată).

4.3 Managementul Oportunității și Punctualității

Oportunitatea rezultatelor statistice reprezintă durata de timp care există între evenimentul sau fenomenul pe care îl descriu și disponibilitatea lor. Informația care le este disponibilă utilizatorilor chiar în perioada în care aceasta rămâne utilă pentru scopurile sale principale, este considerată a fi oportună.

Punctualitatea reprezintă intervalul de timp dintre momentul publicării datelor și momentul planificat pentru publicarea lor, după cum a fost anunțat într-un calendar oficial, stabilit prin regulamente sau printr-un acord anterior dintre parteneri.

Oportunitatea și punctualitatea sunt monitorizate constant de-a lungul timpului și în cadrul programelor, prin rapoarte săptămânale înaintate către managementul superior.

Alegerea unei perioade de timp a unei anchete, inclusiv datele de difuzare reprezintă o decizie a proiectării anchetei, depinzând de obicei de compromisul dintre acuratețe și cost. Există două cataloage separate de rezultate, produse către finele anului și vizibile pe website-ul INS. Acestea specifică datele de difuzare ale publicațiilor electronice și tipărite și a comunicatelor de presă pentru anul următor.

În unele programe, difuzarea unor date preliminare, urmată de publicarea unor cifre revizuite și finale este utilizată ca strategie pentru a oferi datele într-un timp cât mai scurt. În astfel de cazuri, monitorizarea dimensiunii și direcției reviziilor poate ajuta la evaluarea compromisului dintre oportunitate-acuratețe. Aceasta reprezintă, de asemenea, o premisă pentru recunoașterea oricăror abateri persistente sau previzibile ale datelor preliminare, care ar putea fi înlăturate prin estimare.

Se așteaptă îmbunătățiri în ceea ce privește oportunitatea, pe măsura dezvoltării noilor tehnologii și modificării utilizării datelor.

4.4 Managementul Accesibilității

Accesibilitatea rezultatelor statistice reprezintă gradul de ușurință cu care utilizatorii pot obține datele. Ea este determinată de condițiile fizice prin care utilizatorii obțin datele: unde să mergem, cum să comandăm, timpul de livrare, politica de preț, condițiile de piață (drepturile de autor etc.), disponibilitatea datelor micro sau macro, formatele diferite (hârtie, fișiere, CD-ROM, Internet, etc.)

La începutul unei noi anchete sau după o replanificare majoră a unei anchete, conducătorul de anchetă propune mijloacele de diseminare. Aceasta include datele ce trebuie diseminate, modul și momentul diseminării pe baza cererilor cunoscute ale utilizatorilor și a discuțiilor viitoare cu utilizatorii. Propunerea este revizuită de managementul superior, este modificată în conformitate cu deciziile acestora și apoi implementată de Direcția de diseminare.

Principalele mijloace de diseminare ale INS sunt: publicațiile pe hârtie, publicațiile electronice, baza de date online TEMPO, biblioteca publică, biroul de diseminare.

Managerii de program au la dispoziție mai multe opțiuni pentru ca datele să fie mai accesibile în scopuri analitice, printre care:

- Producerea de fișiere de microdate pentru uzul public;
- Oferirea unui serviciu de furnizare personalizată a datelor;
- Contracte cu cercetători externi.

4.5 Managementul Clarității

Claritatea rezultatelor statistice reprezintă gradul de ușurință cu care utilizatorii pot înțelege datele. Ea este determinată de mediul informațional în care sunt prezentate datele, dacă acestea sunt însoțite de metadate corespunzătoare, dacă sunt folosite ilustrații, cum ar fi diagrame și hărți, dacă sunt disponibile informații legate de acuratețea datelor (incluzând toate limitele utilizării) și măsura în care producătorul furnizează asistență suplimentară.

Metadatele

Oferirea unor informații suficiente pentru a permite utilizatorilor să interpreteze corespunzător informațiile statistice face parte din responsabilitățile INS și se referă în primul rând la furnizarea metadatelor adecvate. Informațiile necesare pentru înțelegerea datelor statistice se încadrează în trei mari grupe:

- Conceptele, variabilele și clasificările care stau la baza datelor (metadatele conceptuale);
- Metodologia folosită pentru colectarea și calcularea datelor (metadatele procedurale);
- Indicatori ai acurateții datelor (metadatele operaționale).

Comentariile care însoțesc rezultatele datelor reprezintă un sprijin suplimentar pentru utilizatori.

Mijlocul primar pentru diseminarea metadatelor este Baza de date a metadatelor, disponibilă pe Internet. În prezent nu este conectată cu baza de date TEMPO, la nivelul datelor, acest lucru urmând să fie îmbunătățit. Există de asemenea, mai multe surse separate interne de metadate care trebuie consolidate în Baza de date a metadatelor pentru a-i îmbunătăți conținutul și pentru a împiedica variațiile.

Managementul interpretărilor eronate

Mass-media este monitorizată zilnic, iar, în cazul interpretărilor eronate, Biroul de presă ia măsuri. Interpretările eronate ale datelor de către mass-media sau în zona publicului sunt discutate imediat în mod public de către INS.

4.6 Managementul Coerenței și Comparabilității

Coerența dintre două sau mai multe rezultate statistice se referă la gradul în care procesele statistice care le-au generat au utilizat aceleași concepte – clasificări, definiții și grupuri-țintă – și metode armonizate. Rezultatele statistice coerente pot fi combinate în mod valabil și folosite împreună. **Comparabilitatea** este un caz special al coerenței și se referă la cazul în care rezultatele statistice privesc aceleași date, iar scopul combinării lor este efectuarea unor comparații în timp, între zone geografice sau în alte domenii.

Trei abordări complementare sunt folosite pentru managementul coerenței.

Prima abordare este dezvoltarea și utilizarea conceptelor, definițiilor și sistemelor de clasificare standard pentru toți indicatorii principali, precum și implementarea cadrelor standard internaționale, de exemplu, Sistemul Conturilor Naționale, acolo unde acestea există.

A doua abordare constă în asigurarea că metodele de măsurare nu introduc inconsistențe, de exemplu prin utilizarea:

- Unui registru al întreprinderilor, drept cadru de referință pentru toate anchetele în întreprinderi;
- Registrului exploatațiilor agricole, drept un cadru de referință pentru anchetele în agricultură;
- Eșantionului master al gospodăriilor, pentru anchetele în gospodării;
- Metodologiilor și sistemelor armonizate;
- Ghidurilor calității și standardelor internaționale.

Domeniile pentru îmbunătățire includ documentarea și utilizarea întrebărilor comune, documentarea și utilizarea standardelor de proiectare și formatare standard a chestionarelor.

A treia abordare presupune chiar analiza datelor și compararea și integrarea datelor din surse diferite sau în timp. Scopul este recunoașterea situațiilor în care variația sau inconsistența depășește nivelurile pe care le presupune acuratețea așteptată a datelor. Integrarea datelor în conturile naționale este un exemplu în acest sens. Măsurile privind coerența sunt de asemenea, cerute în anumite rapoarte de calitate elaborate pentru Eurostat în conformitate cu regulamentele UE. De exemplu, un raport privind Ancheta Costului Forței de Muncă prezintă coerența rezultatelor anchetei cu Conturile Naționale, cu Ancheta Forței de Muncă și cu Ancheta Structurală în Întreprinderi.

Și utilizatorii reprezintă o sursă importantă de informații în ceea ce privește coerența, sau mai degrabă lipsa acesteia.

4.7 Managementul Raportării Calității

Scopul raportării calității

Cadrul actual și propus al metodelor de evaluare a calității poate fi descris prin trei niveluri, indicate în Figura 1.

Nivelul 1: Indicatorii calității rezultatului, calitatea procesului și indicatorii de performanță, rapoartele de calitate și feedback-ul utilizatorului/rezultatele anchetei.

Datele obținute din măsurătorile proceselor și rezultatelor sunt selectate și structurate în indicatori de calitate și performanță, care la rândul lor vor fi calculați în rapoartele de calitate. Pentru a măsura percepțiile acestora privind calitatea produsului, se realizează anchete privind gradul de satisfacție al utilizatorilor. Rapoartele calității sunt pregătite în conformitate cu cerințele legislative sau cu cerințele de informații specifice.

Nivelul 2: Auto-evaluări; rapoarte cuprinzătoare și evaluări (peer review)

Pe baza informației și a analizei suplimentare de la Nivelul 1, anchetele și produsele acestora vor fi evaluate, prin comparare, cu un standard intern cuprinzător dezvoltat de Unitatea de calitate. Aceste evaluări vor conferi o imagine mai cuprinzătoare decât evaluările de la Nivelul 1 și, fiind efectuate potrivit unui format comun, ele vor permite unele comparații între anchete.

Nivelul 3: Etichetarea și certificarea

Acest nivel trebuie implementat. Etichetarea are ca obiectiv furnizarea de informații foarte condensate de calitate privind statisticile furnizate în beneficiul utilizatorilor. Va presupune atribuirea unei etichete rezultatelor acelor anchete care satisfac anumite cerințe INS privind calitatea. Certificarea este similară etichetării, dar se face conform unui standard extern precum Modelul de Excelență EFQM și se referă mai degrabă la INS ca un întreg, decât la produse specifice. În prezent nu există planuri pentru etichetare sau certificare.

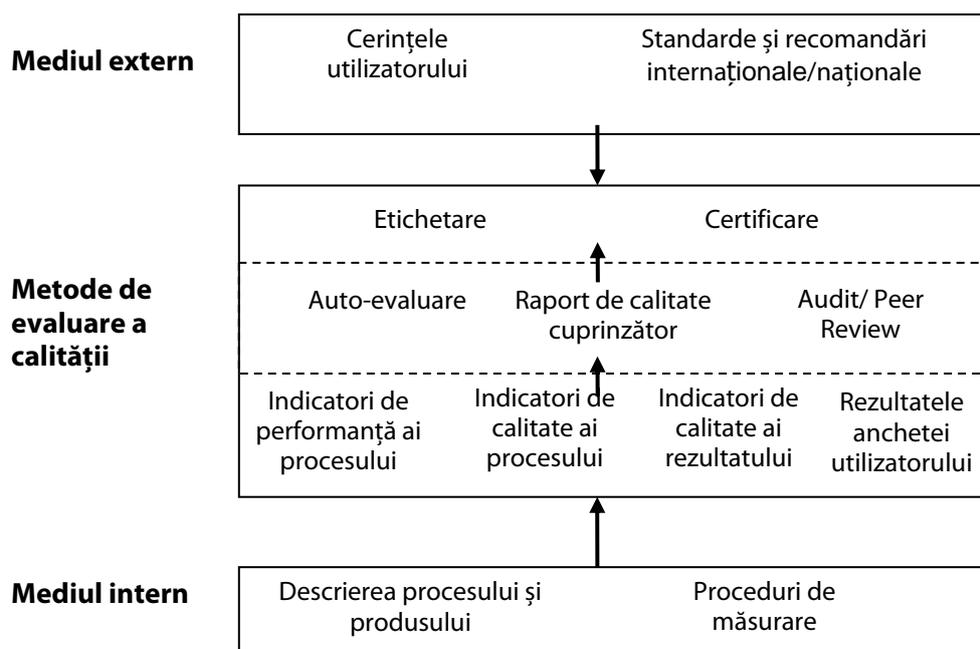


Figura 1: Metode de evaluare a calității

Rapoarte de calitate orientate către utilizator

Așa cum s-a indicat anterior, programul rapoartelor de calitate orientate către utilizator va începe în 2010. În principiu, fiecare anchetă va trebui să pregătească un raport în fiecare an. Totuși, în primul an, vor fi excluse acele anchete care completează rapoarte de calitate pentru Eurostat. Rapoartele vor include setul standard SSE de indicatori de calitate și performanță.

A fost conceput un model ("standard") de raport împreună cu instrucțiunile de completare.

Liste de control (checklist)

INS promovează utilizarea *Listei de control SSE pentru managerii de cercetări statistice (DESAP)* ca mod de evaluare rapidă, dar sistematică a calității unei anchete și ca mijloc de identificare a potențialelor îmbunătățiri.

4.8 Managementul Cadrului Instituțional

Managementul calității se desfășoară în cadrul mediului instituțional al INS. Măsurile concepute pentru a crea și sprijini un mediu și o cultură, care să recunoască importanța calității, includ:

- Un program de recrutare și dezvoltare pentru principalele grupuri ocupaționale;
- Un cadru general pentru instruire și dezvoltare;
- Centrul Național de Pregătire în Statistică;
- Ancheta privind opinia personalului a fost realizată în 2004, iar o alta este programată în 2011;
- Stabilirea zonelor specifice de expertiză și a comisiilor de supraveghere, de exemplu în eșantionare, estimare și în aplicațiile IT;
- Protocoalele și convențiile cu alte departamente și agenții de stat care sunt fie furnizori de date administrative, fie utilizatori de date statistice, sau ambele;
- Proceduri de gestionare a reclamațiilor respondenților;
- Stabilirea metodelor de măsurare a sarcinii de răspuns în anchetele în întreprinderi.

Există o arie vastă pentru îmbunătățiri, de exemplu (1) stabilirea și întărirea domeniilor de expertiză metodologică, (2) încurajarea analizei datelor atât pentru o mai bună informare a utilizatorilor prin intermediul comentariilor asupra datelor de ieșire cât și pentru a detecta problemele și deficiențele datelor, (3) extinderea măsurătorilor sarcinii de răspuns în toate anchetele statistice.

Referințe

Eurostat (2002) *Business Methods Your Guide to Business Statistics Methodology*.

Eurostat (2000) *Handbook on Design and Implementation of Business Surveys*.

Fellegi, I. (1996). *Characteristics of an effective statistical system*. International Statistical Review, 64, 165-197

Trewin, D. (2002). *The importance of a quality culture*. Survey Methodology, 28, 125-133.

5 Anexă: Politici privind Calitatea Sistemului Statistic European

5.1 Declarația privind calitatea SSE

Misiunea Sistemului statistic european

“Noi furnizăm Uniunii Europene și lumii informații de înaltă calitate asupra economiei și societății la nivel european, național și regional și punem la dispoziție informații pentru scopuri decizionale, cercetări și dezbateri publice.”

Viziunea Sistemului statistic european

“SSE va fi un leader mondial în servicii de informații statistice și cel mai important furnizor de informații pentru Uniunea Europeană și Statele Membre. Bazându-se pe principiile și metodele științifice, SSE va oferi și îmbunătăți continuu programul statisticii europene armonizate ce constituie o bază esențială pentru procesele democratice și progresul societății.”

Focalizarea asupra utilizatorului

Noi furnizăm utilizatorilor noștri produse și servicii care răspund nevoilor lor. Nevoile clare și cele neformulate clar, cererile și așteptările utilizatorilor externi și interni vor ghida SSE, membrii săi, angajații și operațiunile.

Îmbunătățirea continuă

Nevoile și cererile utilizatorilor se vor modifica, ca și mediul în care operăm. Globalizarea și progresele metodelor și tehnologiei vor oferi noi posibilități. Este esențial să facem tot posibilul să ne îmbunătățim metodele de lucru pentru a ne folosi de noile posibilități și pentru a răspunde mai bine cererilor utilizatorilor noștri.

Angajamentul privind calitatea produsului

Noi producem informație statistică de înaltă calitate potrivit metodelor științifice în conformitate cu obiectivitatea și confidențialitatea. Furnizăm informații despre caracteristicile principale ale calității fiecărui produs, astfel încât utilizatorii să poată evalua calitatea produsului.

Accesibilitatea informației

Noi furnizăm rezultatele statistice într-o formă accesibilă și prietenoasă utilizatorului. Utilizând posibilitățile noilor mijloace, asigurăm accesul facil la informație. Pe cât posibil, vom crește conștientizarea utilizatorului în ceea ce privește punctele forte și limitările statisticilor produse. Consultarea privind modul în care se utilizează datele este o parte integrantă a diseminării.

Parteneriatul în cadrul Sistemului Statistic European și în afara sa

Cooperarea între membrii actuali și cei viitori ai SSE, precum și cu alte organizații va fi încurajată. Doar lucrând împreună putem învăța de la alții și treptat ne putem dezvolta propriul sistem. Cunoașterea largă a personalului și utilizatorilor noștri, furnizorilor, partenerilor și ale altor părți interesate trebuie combinată pentru a ne permite să excelăm în îndeplinirea scopurilor noastre.

Respectul pentru nevoile furnizorilor de date

Furnizorii de date statistice –respondenții– reprezintă un grup de o importanță specială cu care trebuie întemeiat un parteneriat care să aducă beneficii de ambele părți. Producătorii de

statistici ar trebui să încerce întotdeauna să minimizeze sarcina de răspuns, atât sarcina obiectivă, cât și cea percepută.

Angajamentul leadershipului

Liderii organizațiilor din SSE exercită o conducere personală, activă și vizibilă pentru a crea și susține o cultură a calității. Asigurând o direcție generală clară, prioritizând activitățile de îmbunătățire și stimulând delegarea și inovația, liderii permit personalului să realizeze cu succes activitatea și să încerce continuu să o îmbunătățească.

Managementul sistematic al calității

Noi identificăm sistematic și cu regularitate punctele forte și pe cele slabe în toate domeniile relevante pentru a identifica în mod continuu și pentru a aduce îmbunătățiri acolo unde este nevoie. Pentru dezvoltarea SSE este vitală o orientare strategică pe termen lung. În toate situațiile, efectele pe termen lung trebuie luate în considerare odată cu efectele pe termen scurt, mai evidente.

Procesele eficiente și eficace

Activitățile SSE trebuie văzute ca procese care creează valoare pentru utilizatori. Noi lucrăm eficient pentru a produce rezultate cu cât mai puține resurse posibil și pentru a preveni erorile în procese și produse. Procesele și calitatea lor sunt continuu revizuite și îmbunătățite.

Satisfacția și dezvoltarea personalului

Pentru a atrage și menține un personal competent, satisfacerea nevoilor acestuia este vitală. Membrii SSE trebuie să își trateze angajații ca pe niște resurse cheie.

5.2 Codul de Practici al Statisticilor Europene

Pentru autoritățile naționale și comunitare de statistică

Adoptat de Comitetul pentru program statistic la data de 24 februarie 2005 și promulgată în Recomandarea Comisiei din 25 mai 2005 privind Independența, Integritatea și Responsabilitatea Autorităților Statistice Naționale și Comunitare

Codul de Practici al Statisticilor Europene se bazează pe 15 principii. Autoritățile guvernamentale și cele statistice din Uniunea Europeană se angajează să adere la principiile stabilite în acest Cod, acoperind mediul instituțional, procesele și rezultatele statistice. Un set de indicatori de bună practică pentru fiecare dintre cele 15 principii asigură o referință pentru revizuirea implementării Codului.

Mediul instituțional

Factorii instituționali și organizatorici au o influență considerabilă asupra eficienței și credibilității unei autorități statistice care produce și diseminează statistici europene. Problemele relevante sunt independența profesională, autoritate în materie de culegere a datelor, alocarea proporțională a resurselor, angajamentul vis-a-vis de calitate în statistică, confidențialitatea statistică, imparțialitatea și obiectivitatea.

PRINCIPIUL 1: INDEPENDENȚA PROFESIONALĂ

Independența profesională a autorităților statistice față de alte departamente și organisme legislative, administrative sau având competențe în formularea de politici, precum și față de operatorii din sectorul privat asigură credibilitatea statisticilor europene.

Indicatori

- Independența autorităților statistice față de orice tip de intervenții externe, politice sau de altă natură, în producerea și diseminarea statisticilor oficiale este stipulată în lege.
- Șeful autorității statistice are o poziție ierarhică suficient de importantă pentru a se asigura contactul la nivel înalt cu autoritățile responsabile cu formularea de politici și organismele publice administrative. Calificarea lui/ei profesională trebuie să fie de cel mai înalt nivel.
- Șeful autorității statistice este cel care asigură o producție și o diseminare independentă a statisticilor europene.
- Șeful autorității statistice este singurul responsabil cu decizia asupra metodelor, standardelor și procedurilor statistice, precum și asupra conținutului și momentelor publicărilor de date statistice.
- Programele de lucru statistice sunt publicate și rapoarte periodice descriu progresele realizate.
- Publicările de date statistice se disting clar și au loc separat de declarațiile cu caracter politic sau privitoare la formularea de politici.
- Autoritatea statistică, când este cazul, face comentarii publice asupra informațiilor statistice, inclusiv despre criticile și lipsurile statisticilor oficiale

PRINCIPIUL 2: AUTORITATE ÎN MATERIE DE CULEGERE A DATELOR

Autoritățile statistice trebuie să dețină o împuternicire legală clară pentru culegerea informațiilor în vederea elaborării de statistici europene. La cererea autorităților statistice, administrațiile, întreprinderile și gospodăriile, precum și populația în general, pot fi obligate prin lege să permită accesul la date sau să furnizeze date pentru elaborarea statisticilor europene.

Indicatori

- Împuternicirea pentru culegerea informațiilor în vederea producerii și diseminării de statistici oficiale este stipulată în lege.
- Legislația națională permite autorității statistice să folosească registrele administrative în scopuri statistice.
- Pe baza unui act juridic, autoritatea statistică poate pretinde obligativitatea de a răspunde la anchetele statistice

PRINCIPIUL 3: ADECVAREA RESURSELOR

Resursele aflate la dispoziția autorităților statistice trebuie să fie suficiente pentru a răspunde cerințelor statisticilor europene.

Indicatori

- Se dispune de resurse umane, financiare și informatice, adecvate atât din punct de vedere cantitativ cât și calitativ, pentru a răspunde cerințelor statisticilor europene.
- Sfera de cuprindere, gradul de detaliu și costul statisticilor europene sunt proporționale cu nevoile existente.
- Există proceduri pentru evaluarea și justificarea cererilor de noi statistici europene în raport cu costul lor.
- Există proceduri pentru evaluarea nevoii continue de statistici europene, cu scopul de a vedea dacă producția vreuneia dintre ele poate fi stopată sau redusă pentru a disponibiliza resurse

PRINCIPIUL 4: ANGAJAMENTUL PRIVIND CALITATEA

Toți membrii SSE se angajează să lucreze și să coopereze în conformitate cu principiile stabilite în Declarația privind calitatea Sistemului Statistic European.

Indicatori

- Calitatea produselor este monitorizată regulat în conformitate cu componentele calitative ale SSE.
- Există procese pentru a monitoriza calitatea culegerii, prelucrării și diseminării de statistici.
- Există procese pentru abordarea considerentelor referitoare la calitate, inclusiv eventuale soluții de compromis în domeniul calității, precum și pentru orientarea planificării anchetelor existente și viitoare.
- Recomandările privind calitatea în statistică sunt documentate și personalul este bine instruit. Aceste recomandări sunt specificate în scris și aduse la cunoștința publicului.
- Are loc o revizuire regulată și amănunțită a produselor statistice cheie, folosindu-se auditori externi, dacă este cazul

PRINCIPIUL 5: CONFIDENȚIALITATEA STATISTICĂ

Confidențialitatea informațiilor pe care le furnizează respondenții (gospodării, întreprinderi, administrații, etc.) și utilizarea lor numai în scopuri statistice trebuie să fie pe deplin garantate.

Indicatori

- Confidențialitatea statistică este garantată prin lege.
- Personalul autorității statistice semnează angajamente juridice de confidențialitate la numirea în funcție.
- Penalizări substanțiale sunt prevăzute pentru orice încălcare deliberată a confidențialității statistice.
- Sunt furnizate instrucțiuni și recomandări privind protecția confidențialității statistice în procesele de producție și diseminare. Aceste recomandări sunt specificate în scris și aduse la cunoștința opiniei publice.
- Sunt în vigoare prevederi de natură fizică și tehnologică pentru a proteja securitatea și integritatea bazelor de date statistice.
- Protocoale stricte se aplică utilizatorilor externi care accesează microbazele de date statistice în scopuri de cercetare

PRINCIPIUL 6: IMPARȚIALITATEA ȘI OBIECTIVITATEA

Autoritățile statistice trebuie să producă și să disemineze statistici europene respectând independența științifică și într-o manieră obiectivă, profesionistă și transparentă prin care toți utilizatorii sunt tratați echitabil.

Indicatori

- Statisticile sunt compilate în conformitate cu criteriile obiective determinate de considerente științifice.
- Alegerea surselor și tehnicilor statistice se bazează pe considerente de tehnică statistică.
- Erorile detectate în statisticile publicate sunt corectate cât de repede posibil și sunt făcute publice.
- Informațiile privind metodele și procedurile utilizate de autoritatea statistică se află la dispoziția publicului.
- Datele și termenele publicărilor de date statistice sunt anunțate în avans.
- Toți utilizatorii au acces egal și simultan la publicările de date statistice și orice acces privilegiat anterior publicării acordat utilizatorilor externi este limitat, controlat și făcut public. În cazul în care se produc scurgeri de informații, dispozițiile privind accesul anterior publicării trebuie revizuite pentru a garanta imparțialitatea.
- Publicările de date statistice și declarațiile prezentate la conferințele de presă sunt obiective și imparțiale

Procese statistice

Standardele, recomandările și bunele practici europene și internaționale trebuie respectate întocmai în procesele utilizate de autoritățile statistice pentru a organiza, culege, prelucra și disemina statistici oficiale. Credibilitatea statisticilor este sporită de reputația unui bun management și a eficienței. Aspectele relevante sunt metodologia solidă, procedurile statistice adecvate, sarcina de răspuns nu excesivă pentru respondenți și eficiența sub raportul costurilor.

PRINCIPIUL 7: METODOLOGIE SOLIDĂ

Statisticile de calitate trebuie să se bazeze pe o metodologie solidă. Aceasta necesită instrumente, proceduri și cunoștințe de specialitate adecvate.

Indicatori

- Cadrul metodologic general al autorității statistice urmează standarde, recomandări și bune practici europene și internaționale.
- Există proceduri pentru a garanta aplicarea coerentă a conceptelor, definițiilor și clasificărilor standard la toate nivelurile autorității statistice.
- Registrul întreprinderilor și cadrul pentru anchetele populației sunt evaluate regulat și ajustate dacă este necesar în vederea asigurării unei înalte calități.
- Există o concordanță detaliată între clasificările naționale și sistemele sectoriale, pe de o parte, și sistemele europene care le corespund, pe de altă parte.
- Sunt angajați absolvenți ai disciplinelor academice relevante.
- Personalul asistă la cursuri de instruire și conferințe internaționale relevante și ține legătura cu colegi statisticieni la nivel internațional pentru a învăța de la cei mai buni și pentru a-și îmbunătăți cunoștințele de specialitate.
- Se organizează activități de cooperare cu comunitatea științifică pentru îmbunătățirea metodologiei și, atunci când este posibil, se recurge la revizuirii externe pentru a evalua calitatea și eficiența metodelor implementate și a promova instrumente mai bune

PRINCIPIUL 8: PROCEDURI STATISTICE ADECVATE

Statisticile de calitate trebuie să se bazeze pe proceduri statistice adecvate, implementate de la culegerea datelor până la validarea acestora.

Indicatori

- Când statisticile europene se bazează pe date administrative, definițiile și conceptele utilizate în scop administrativ trebuie să constituie o bună aproximare a celor solicitate în scopuri statistice.
- În cazul anchetelor statistice, chestionarele sunt sistematic testate înainte de culegerea datelor.
- Proiectarea anchetelor, selectarea eșantioanelor și ponderile eșantioanelor au o bază solidă și sunt revizuite sau actualizate regulat când se consideră necesar.
- Operațiunile de teren, introducerea datelor și codificarea sunt constant monitorizate și revizuite conform nevoilor existente.
- Se utilizează sisteme informatice de identificare și corectare automată a greșelilor, respectiv de imputare, ele fiind revizuite sau actualizate regulat, conform cerințelor.
- Revizuirile urmează proceduri standard, bine stabilite și transparente

PRINCIPIUL 9: SARCINA ADECVATĂ PENTRU RESPONDENȚI

Sarcina de raportare trebuie să fie proporțională cu nevoile utilizatorilor și să nu se dovedească excesivă pentru respondenți. Autoritatea statistică monitorizează sarcina de răspuns și stabilește ținte pentru reducerea sa în timp.

Indicatori

- Sfera de cuprindere și gradul de detaliu al cerințelor statisticilor europene se limitează la ceea ce este absolut necesar.
- Sarcina de raportare este împărțită cât de mult posibil între populațiile care fac obiectul anchetei prin tehnici de eșantionare adecvate.
- Informația solicitată întreprinderilor este, pe cât posibil, prompt disponibilă în conturile acestora și, în măsura posibilităților, se utilizează mijloace externe pentru a facilita transmiterea ei.
- Cele mai bune estimări și aproximări sunt acceptate atunci când nu sunt prompt disponibile detalii exacte.
- Se folosesc surse administrative de câte ori este posibil pentru a evita duplicarea cerințelor de informații.
- Împărțirea datelor la nivelul autorităților statistice este o practică generalizată în scopul de a evita multiplicarea anchetelor

PRINCIPIUL 10: EFICIENȚA SUB RAPORTUL COSTURILOR

Resursele trebuie utilizate în mod eficient.

Indicatori

- Măsurile interne și independente externe monitorizează utilizarea resurselor de către autoritatea statistică.
- Operațiunile administrative obișnuite (de ex. extragerea, codificarea și validarea datelor) sunt automate în măsura posibilului.
- Potențialul de productivitate al tehnologiei informației și comunicațiilor este optimizat pentru culegerea, prelucrarea și diseminarea datelor.
- Se fac eforturi proactive pentru a îmbunătăți potențialul statistic al registrelor administrative și a evita anchete directe costisitoare

Rezultatul statistic

Statisticile disponibile trebuie să îndeplinească nevoile utilizatorilor. Statisticile sunt în conformitate cu standardele calitative europene și servesc nevoilor instituțiilor europene, guvernelor, instituțiilor de cercetare, intereselor comerciale și publicului în general. Problemele importante privesc măsura în care statisticile sunt relevante, exacte și fiabile, prompte, coerente, comparabile la nivelul regiunilor și țărilor și rapid accesibile utilizatorilor.

PRINCIPIUL 11: RELEVANȚA

Statisticile europene trebuie să îndeplinească nevoile utilizatorilor.

Indicatori

- Funcționează procese de consultare a utilizatorilor, de monitorizare a relevanței și utilității practice a statisticilor existente în ceea ce privește îndeplinirea nevoilor acestora și de consiliere privind noile lor nevoi și priorități.
- Nevoile prioritare sunt îndeplinite și se reflectă în programul de lucru.
- Anchete privind satisfacția utilizatorilor sunt efectuate periodic

PRINCIPIUL 12: PRECIZIA ȘI ÎNCREDEREA

Statisticile europene trebuie să reflecte realitatea cu precizie și fiabilitate.

Indicatori

- Sursele de date, rezultatele intermediare și produsele statistice sunt evaluate și validate.
- Erorile de eșantionare și erorile de altă origine decât eșantionarea sunt măsurate și documentate sistematic în conformitate cu cadrul componentelor calitative ale SSE.
- Studii și analize ale revizuirilor sunt efectuate constant și utilizate intern, ele stând la baza proceselor statistice

PRINCIPIUL 13: OPORTUNITATEA ȘI PUNCTUALITATEA

Statisticile europene trebuie să fie diseminate într-o manieră promptă și punctuală.

Indicatori

- Promptitudinea (respectarea termenelor) urmează standarde de diseminare europene și alte standarde internaționale.
- Se fixează un termen zilnic standard pentru publicarea statisticilor europene.
- Periodicitatea statisticilor europene are în vedere cerințele utilizatorilor pe cât de mult posibil.
- Orice neconcordanță în ceea ce privește calendarul de diseminare este făcută publică în avans, explicată și este stabilită o nouă dată a publicării.
- Rezultatele preliminare de o calitate agregată acceptabilă pot fi diseminate atunci când acest lucru este considerat util

PRINCIPIUL 14: COERENȚA ȘI COMPARABILITATEA

Statisticile europene trebuie să aibă coerență internă și în timp și să fie comparabile între regiuni și țări; trebuie să fie posibile combinarea și folosirea comună a datelor conexe din diferite surse.

Indicatori

- Statisticile au coerență și concordanță internă (de ex.: se respectă identitățile aritmetice și contabile).
- Statisticile sunt coerente sau reconciliabile pe o perioadă de timp rezonabilă.
- Statisticile sunt compilate pe baza unor standarde comune în ceea ce privește sfera de cuprindere, definițiile, unitățile și clasificările din diferite anchete și surse.
- Statisticile provenite din anchete și surse diferite sunt comparate și reconciliate.
- Comparabilitatea transnațională a datelor este periodic analizată în studii metodologice, efectuate în strânsă cooperare cu statele membre și Eurostat.

PRINCIPIUL 15: ACCESIBILITATEA ȘI CLARITATEA

Statisticile europene trebuie să fie prezentate într-o formă clară și ușor de înțeles, să fie diseminate într-o manieră adecvată și convenabilă, să fie disponibile și accesibile respectându-se imparțialitatea și să fie însoțite de metadata și recomandări.

Indicatori

- Statisticile sunt prezentate într-o formă care facilitează interpretarea adecvată și comparațiile semnificative.
- Serviciile de diseminare utilizează tehnologia modernă a informației și comunicațiilor precum și tradiționalul suport de hârtie.
- Analize personalizate sunt furnizate atunci când acest lucru este realizabil și sunt făcute publice.
- Accesul la microdate este permis în scopuri de cercetare atunci când este posibil. Acest acces este supus unor protocoale stricte.
- Metadatale sunt documentate în conformitate cu sisteme de metadata standardizate.
- Utilizatorii sunt informați în legătură cu metodologia proceselor statistice și calitatea producției statistice raportat la criteriile calitative ale SSE.

1 Introduction

Purpose of Guidelines

This document brings together the many issues that need to be considered in pursuing quality objectives at the National Institute of Statistics (NIS). It focuses on how to assure quality through effective and appropriate design and implementation of surveys and other statistical processes throughout their entire life cycles. Its specific objectives are to promulgate quality concepts and promote quality awareness, to provide technical guidance on survey design and conduct, and to summarise the organisation's quality assurance framework.

In this context the terms *survey* and *statistical process* include sample survey, census, survey based on administrative data, production of price or other economic index, or any other statistical compilation such as the National Accounts and Balance of Payments, as further discussed below.

Users and Uses

The Guidelines are primarily intended for use by staff engaged in the design and/or conduct sample surveys, censuses and other statistical processes. They may also serve senior managers and users who are seeking assurance that survey methods are appropriate and output data quality is likely to be good.

Terminology

To the extent possible, the terminology is in line with the European Statistical System (ESS) Glossary. Where a term is not in the ESS Glossary its definition is drawn from another international source, where available, such as the Metadata Common Vocabulary (MCV) created by the OECD and other international organizations working on the Statistical Data and Metadata Exchange (SDMX) initiative. The NIS Statistical Glossary is also taken into account.

Types of Survey

From a technical statistical perspective, the methods of producing NIS statistics show a great diversity. Hence a single set of recommendations, especially those regarding accuracy, cannot apply to all production processes, it is necessary to introduce some distinctions. For the purpose of this document, and based on the *ESS Handbook for Quality Reports* (Eurostat, 2009), six types of survey and statistical process are distinguished, as follows.

1. *Sample survey*. This is a survey based on a (usually) probabilistic sampling procedure involving direct collection of data from respondents. For this kind of survey there is an established theory on accuracy that leads to well defined error components (sampling and non-sampling errors). Examples are the Household Labour Force Survey, the Labour Cost Survey, and the Survey of Direct Foreign Investments.
2. *Census*. This can be viewed as a special case of the sample survey in which data are collected from all units in the survey frame. An example is the 2002 Population and Households Census
3. *Survey using administrative source(s)*. This sort of survey makes use of data collected for other purposes than direct production of statistics.
4. *Survey involving multiple data sources*.
5. *Price or other economic index production process*.
6. *Statistical compilation*. Data are assembled from a variety of primary sources to produce aggregate data with a special conceptual significance, for example, the National Accounts and the Balance of Payments.

Content of Guidelines

Chapter 2 aims to set the Guidelines within an international context. It describes the basic quality concepts on which the NIS approach to quality is based, including a review of the ESS and other international quality documentation available and the definitions of output quality and process quality. It is substantially extracted from the *ESS Handbook for Quality Reports* (Eurostat, 2009).

Chapter 3 presents technical guidelines for use in the context of the design, redesign, development or operation of a survey or other statistical process. The guidelines are written with censuses and sample surveys as the main focus but also apply to the other types of statistical process defined above.

Whereas Chapter 3 focuses on individual surveys and other statistical processes, Chapter 4 refers to the whole organisation, describing the quality framework in place.

The Appendices contain copies of two relevant ESS policy documents.

2 Quality Concepts

2.1 International Perspective

Evolution of Quality Management

Modern quality management has its origins *statistical quality control* as designed for production lines in factories in the 1920s and still in practice. Walter Shewhart, sometimes called the “father of statistical quality control” introduced the notion of variation due to assignable and to random causes and the elimination of the former using process quality control charts. The focus of quality control is inspection and correction. Subsequently, in the 1940s and 1950s, more emphasis was placed on preventing defects occurring rather than correcting them, for example feeding back the results of quality inspection as the basis for modifying the problems in the production lines and for employee training. This was referred to as *upstream quality control*, and the broader range of quality measures were referred to as *quality assurance*.

Over the next 30 years, as a result of papers written and advice given by quality gurus such as William Edwards Deming, renowned for his *Fourteen Principles of Management*, and Joseph Juran, who championed the human dimension, the notion of quality assurance was further extended to *total quality management (TQM)*. TQM principles are typically expressed along the following lines.

- *Customer focus*: an organisation depends upon its customers and thus must understand and strive to meet their needs; customers are central in determining what constitutes good quality; quality is what is perceived by customers rather than by the organisation.
- *Leadership and constancy of purpose*: leaders establish unity of purpose and direction of an organisation; they must create and maintain an internal environment that enables staff to be fully involved in achieving the organisation’s objectives; quality improvements require leadership and sustained direction.
- *Involvement of people*: people at all levels are the essence of an organisation; their full involvement enables their abilities to be fully used.
- *Process approach*: managing activities and resources as a process is efficient; any process can be broken down into a chain of sub processes, for which the output of one process is the input to the next.
- *Systems approach to management*: identifying, understanding and managing processes as a system contributes to efficiency and effectiveness.
- *Continual improvement*: continual improvement should be a permanent objective of an organisation.
- *Factual approach to decision making*: effective decisions are based on the analysis of information and data.
- *Mutually beneficial supplier relationships*: an organisation and its suppliers are interdependent and a mutually beneficial relationship enhances both.

ISO 9000 Quality Management System Series

The statement of TQM principles given above was extracted from the ISO 9000 series, which are the most widely used quality standards in the world. There are actually three standards:

- *ISO 9000: 2005 Quality Management Systems - Fundamentals and Vocabulary* describes the fundamentals of quality management systems, including quality principles (as reproduced above), terminology and models;

- *ISO 9001: 2000 Quality Management Systems – Requirements* specifies the required characteristics of a quality management system. It is the best known of the three standards as it is the one with respect to which organisations may seek to be certified; and
- *ISO 9004: 2000 Quality Management Systems – Guidelines* is designed to assist organisations having mature quality management systems to obtain performance improvements.

EFQM Quality Standard

The European Foundation for Quality Management has produced what is, in effect, a quality standard similar to the ISO 9000 series. Whilst originally designed for use by commercial organisations in the European Union, the *EFQM Excellence Model* is now globally recognised with thousands of organizations around the world using the EFQM principles to guide their business strategies and operations. *Excellence* is defined as *outstanding practice in managing the organisation and achieving results* and is said to require leadership commitment to principals along similar lines to the ISO 9000 principles, namely: results orientation; customer focus; leadership and constancy of purpose; management by processes and facts; people development and involvement; continuous learning, innovation and improvement; partnership development; and corporate social responsibility.

ISO 20252:2006 Market, Opinion and Social Research – Vocabulary and Service Requirements

This new international standard establishes terminology and service requirements for organizations conducting market, opinion and social research and sets a common level of quality for market research globally. It has its origins in the British Standards Institute standard BS7911, which it supersedes. It comprises sections on quality management systems, managing the research, data collection (including fieldwork, self-completion, and collection from secondary sources), data management and processing (capture, coding, editing, analysis, security), and reporting.

2.2 Introducing Quality Management at the NIS

Need to Interpret International Quality Standards

With the exception of ISO 20252, the quality standards and models described in the previous section are intended to cover all organizations whatever their organisational structure, processes and products and are expressed in very general terms. Hence, as stated in the introduction to ISO 9001:2000:

- the design and implementation of an organisation's quality management system is influenced by varying needs, particular objectives, the products provided, the processes employed and the size and structure of the organisation. It is not the intent of the standard to imply uniformity in the structure of quality management systems or uniformity of documentation.

Thus, in developing its quality management system (referred to as a *quality management framework (QMF)*), the NIS has to interpret the international standards within its specific context.

NIS Context

The context within which the NIS operates may be characterized as follows.

- It is government not private enterprise and not profit based. It supplies data to non-paying *users* rather than to paying *customers*. Thus, for the most part, the users cannot influence quality through purchase decisions.

- Some of the users are actually *internal users*, for example the national accounts area is a user as well as a producer.
- The primary inputs at national level are typically data from individual enterprises, households and persons, whether collected directly or through administrative processes.
- The core production processes are transformations of these individual data into aggregate data, i.e., statistics, and their assembly into national statistical products.
- The primary products, referred to in this document as *outputs*, are statistics and accompanying services.

Elements of NIS Quality Management Framework

Taking these characteristics into account, the NIS quality management framework (QMF) should comprise:

- Quality policy – a short statement by senior management indicating the extent of its commitment to quality management;
- Quality model – a definition of what is meant by quality, usually elaborated in terms of quality components;
- Quality objectives, standards and guidelines – target quality objectives together with international or local standards and guidelines adopted by the NIS;
- Quality assurance procedures - part of, or embedded in, the production processes to the extent possible;
- Quality assessment procedures – sometimes incorporated in the quality assurance procedures, more frequently conducted on a periodic basis, for example based on a self-assessment checklist such as the DESAP checklist;
- Quality measurement procedures - specifically including a set of quality and performance indicators, with procedures for collecting the corresponding data values being embedded in the production processes to the extent possible;
- Quality improvement programme – continual improvement and re-engineering initiatives.

The aim is to put all these elements in place as further discussed in the following paragraphs and in Chapter 4.

2.3 Quality Policy

International Policies

The UN Statistical Commission provides some guidance on the sort of environment within which quality management can flourish through its *UN Fundamental Principles of Official Statistics* for national statistical systems, promulgated in 1994. There are ten principles.

Within the ESS, the *ESS Quality Declaration* was adopted by the Statistical Programme Committee in 2001 as a formal step towards total quality management in the ESS, in line with the EFQM Excellence Model. It is just one and a half pages long, comprising the ESS mission statement, the ESS vision statement and ten principles which borrow from the UN Fundamental Principles and tailor them in the ESS context.

The *European Statistics Code of Practice (CoP)* was promulgated by the European Commission in 2005. It commits Eurostat and the 27 Member State offices, to fifteen principles covering the *institutional environment, statistical processes and outputs*.

Quality is also a consideration in other regulations adopted by the European Council and Parliament creating the legal basis for the provision of European statistics in various domains. Council Regulations are themselves quality assurance mechanisms, setting specific timeliness

targets establishing methodological standards leading to enhanced accuracy and comparability, and covering relevance in the form of the needs of European institutions for national statistics.

NIS Quality Policy

As coordinator of the Romanian National Statistical System, the NIS mission is to satisfy to the extent possible the information needs of all categories of statistical data users through the collection, production and dissemination of relevant, accurate, coherent, timely and accessible statistical data, required to inform decisions on economic and social development of the country and the realities of Romanian society.

Over the last six years the NIS has made a lot of progress in the direction of total quality management and ensuring a quality culture within the organization. The aim during the next few years is to achieve methodological and operational performances at levels comparable to those in the most advanced national statistical institutes within the EU Member States.

2.4 Quality Model

QMF development depends upon a precise definition of quality. The starting point is the definition of output quality from a customer/user perspective. The most general and succinct definition of product quality is *fitness for use*. In slightly more detail, but rather obscurely, ISO 9000 defines product quality, which in the case of statistics is synonymous with output quality, as:

- the degree to which a set of inherent characteristics fulfils requirements.

A somewhat more comprehensible definition provided by ISO 8402:1986 is:

- the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs.

These broad definitions provide a basic notion of output quality, but need to be accompanied by a more precise interpretation of the *components* of output quality. This is provided in the *ESS Quality Definition*, which was presented at the October 2003 meeting of the ESS Working Group *Assessment of Quality in Statistics*. Although never given a more formal status, this document has provided the basis for defining output quality components in all subsequent ESS quality related documents, and has been adopted by the NIS.

Output Quality Components

The NIS defines output quality in accordance with the *ESS Quality Definition*. In broad terms the quality of the NIS outputs are assessed in terms of their “fitness for use”. More precisely, output quality is measured in terms of six quality components: relevance, accuracy, timeliness and punctuality, clarity and accessibility, and coherence and comparability.

Relevance is the degree to which statistical output meets current and potential user needs. It depends on whether all the statistics required by users are produced and the extent to which concepts used (definitions, classifications etc.) reflect what users need. Users have a variety of, sometimes conflicting, needs. The challenge for NIS is to weigh and balance the needs of current and potential users in designing surveys within the organization’s resource constraints.

The **accuracy** of statistical output in the general statistical sense is the degree of closeness of estimates to the true values. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, response, non-response, and processing).

The **timeliness** of statistical output is the length of time between the event or phenomenon they describe and their availability. Information that is available to users well within the period during which it remains useful for its main purposes is considered to be timely. Decisions regarding timeliness typically involve trade-offs with relevance and accuracy.

The **punctuality** of statistical output is the time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by regulations or previously agreed among partners.

The **accessibility** of statistical output is the measure of the ease with which users can obtain the data. It is determined by the physical conditions by means of which users obtain data: where to go, how to order, delivery time, pricing policy, marketing conditions (copyright, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, Internet, etc.)

The **clarity** of statistical output is the measure of the ease with which users can understand the data. It is determined by the information environment within which the data are presented, whether the data are accompanied with appropriate metadata describing concepts, sources, methods, whether use is made of illustrations such as graphs and maps, whether information on data accuracy are available (including any limitations on use) and the extent to which additional assistance is made available to users.

The **coherence** of two or more statistical outputs refers to the degree to which the statistical processes by which they were generated used the same concepts - classifications, definitions, and target populations – and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly. Examples of joint use are where the statistical outputs refer to the same population, reference period and region but comprise different sets of data items (say, employment data and production data) or where they comprise the same data items (say, employment data) but for different reference periods, regions, or other domains.

Comparability is a special case of coherence and refers to the latter example where the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains.

Process Quality Components

Output quality is achieved through *process quality*, which has two broad aspects:

- *effectiveness*: which leads to the outputs of good quality; and
- *efficiency*: which leads to their production at minimum cost to the NIS and to respondents and other data providers.

The *components* of process quality may be formulated in two groups based on the first ten principles in the ESS Code of Practice, as follows.

Institutional Environment

- *Professional independence*: NIS staff is professionally independent from other policy, regulatory or administrative departments and from private sector organizations.
- *Mandate for data collection*: the NIS has a clear legal mandate to collect information. Providers can be compelled by law to allow access to or to deliver data.
- *Adequacy of resources*: the resources available are sufficient to meet systems and processing requirements.
- *Quality commitment*: staff commits themselves to work and cooperate according to the principles stated in the Quality Declaration of the European statistical system.
- *Statistical confidentiality*: the privacy of data providers (households, enterprises, administrations and other respondents) and the confidentiality of the information they provide, and its use only for statistical purposes, is guaranteed.

- *Impartiality and objectivity*: production and dissemination of statistics respect scientific independence and are conducted in an objective, professional and transparent manner in which all users are treated equitably.

Surveys and Statistical Infrastructure

- *Sound methodology*: good tools, procedures and expertise.
- *Appropriate statistical procedures*: are implemented throughout the survey life cycle.
- *Non-excessive burden on respondents*: reporting burden is in proportion to the needs of the users and not excessive. It is monitored over time and targets are set for its reduction.
- *Cost effectiveness*: resources are effectively used.

Balancing Quality Considerations

The output and process quality components are interrelated. There is no general model that brings them together so as to optimize or to prescribe a level of quality. Achieving an acceptable level of quality is the result of addressing and balancing all quality components over time. This balance is a critical aspect of the design, implementation and operation of NIS surveys.

The quality guidelines presented in Chapter 3 are an aid in promoting good statistical practices, in particular sound methodology, appropriate statistical procedures, non-excessive burden on respondents, cost effectiveness and all aspects of output quality.

2.5 Quality Assessment and Reporting

ESS Methods and Procedures

- The *ESS Standard for Quality Reports*, which was updated in 2009, provides recommendations for the preparation of comprehensive quality reports by NSOs and Eurostat units.
- The *ESS Handbook for Quality Reports* also updated in 2009 provides more details and examples.
- Both documents contain the most recent version of *Standard Quality Indicators* for use in summarizing the quality of statistical outputs.
- The *ESS Quality Glossary* covers many technical terms in ESS quality documentation, providing a short definition of each term and indicating the source of the definition.
- A more comprehensive and up to date glossary is the *Metadata Common Vocabulary*, developed by a partnership of international organisations including Eurostat.
- The *European Self-Assessment Checklist for Survey Managers* (DESAP) enables the conduct of quick but systematic and comprehensive quality assessments of a survey and its outputs and identification of potential improvements.
- There is a *Condensed Version of Checklist for survey managers* (DESAP) that contains only selected key questions, and there is an *Electronic Version of Self Assessment Checklist* with an *Electronic Version User Guide*.
- The *Handbook on Data Quality - Assessment Methods and Tools* (DatQAM) details the full range of methods for assessing process and output quality and the tools that support them.
- The *ESS Handbook on Improving Quality by Analysis of Process Variables* describes a general approach and useful tools for identifying, measuring and analysing key process variables.

(The latest versions of all these documents are available from the quality pages on the Eurostat website.) Taking into account the approaches presented, the NIS quality reporting arrangements are discussed in Chapter 4.

3 Quality Guidelines

Introductory Remarks

These guidelines are intended to help in designing a new survey or redeveloping or enhancing an existing one. They are organized in sections, most of which correspond to the main sub-processes (steps) of which a survey is typically comprised. This seems to be a more natural structure for survey managers (the main users) than organizing the same information according to process and output quality components. Consideration was given to using the same structure as the recently promulgated *Generic Statistical Business Process Model*. However, this model has its origins in an IT environment and includes a sharp distinction between design and build functions, whereas from a survey manager's perspective it is more appropriate to say all there is to say about a particular sub-process, say sampling, in a single section, whether it is related to designing, building or operating.

Section 3.1 provides guidelines to assist in identifying objectives, uses and users. The following sections cover subsequent survey sub-processes in roughly chronological order. However, some sub-processes may take place concurrently, for example, frame development, sampling, and questionnaire design. There are significant interdependencies between some sub-processes, for example, between questionnaire design and data collection and capture. Other sections, such as documentation, refer to activities that do not constitute distinct sub-processes on their own, but rather take place within the context of several sub-processes. Finally, aspects relating to administrative data use are discussed in a separate section.

All sections have the same four part structure, as follows.

- *Scope and Purpose*, comprising a definition and description of the sub-process or activity, and an indication of its potential impact on quality, thus establishing the context for the following guidelines.
- *Principles*, comprising broad, underlying policies and approaches that govern the design of the sub-process or activity, with an emphasis on those that relate to quality.
- *Guidelines*, comprising known good practices and providing a comprehensive technical checklist to aid survey design and operations. Judgement is needed in deciding practices are appropriate in any particular case and when and how they should be implemented.
- *References*. The existence and relevance of any relevant NIS or international policy or standard is indicated together with documents that provide more details in specific aspects.

3.1 Objectives, Uses and Users

Scope and purpose

An initial planning exercise must be conducted to establish the basic data needs and the overall feasibility of a proposed new or redesigned survey before it is approved. This includes definition of the broad objectives, the user population being targeted, and the key questions or issues to which analysis based on the data will be directed. Following approval, the results of the initial planning must now further elaborated to ensure that the new or redesigned activity will meet specific user requirements.

Objectives are the purposes for which information is required, stated within the context of the program, research problem or hypotheses that gives rise to the need for data.

Specific *uses* indicate more precisely the data needs, for example, by describing what decisions may be made based on the data collected and how the data will support these decisions. For repeated surveys, additional uses may evolve over time.

Users are the organizations and individuals expected to use the data.

Arriving at a consensus on specific objectives, uses and users is essential for rational decisions about survey design.

Principles

Specification of objectives and uses in consultation with potential users leads to the development of a detailed plan for survey. Users can help in describing more precisely the survey objectives. Survey developers can identify the conceptual, methodological and operational issues that they must resolve and can suggest a reasonable schedule.

It is important to have a clear understanding and to formulate a concrete statement of the objectives in terms of hypotheses to be tested and/or specific data requirements, including the quality expected, budget constraints and expected delivery dates.

A statement of objectives is an important means of communicating to users what outputs can be expected and their limitations. It also provides subsequent potential users, who may have different objectives, with the means to assess the extent to which the outputs may meet their specific needs.

Guidelines

- Focus analysis of user needs on finding the most cost-effective solutions for both the short and long term. Before embarking on a design/redesign, analyse currently available statistics in the area in terms of sources, frequency, quality, timeliness, etc. Consider the trade-off between limitations of currently available data and the cost and time required to produce additional data.
- Develop survey objectives in partnership with important users and stakeholders. Identify important users and groups of users, for example policy makers. Include potential users within the NIS itself, for example the National Accounts. Establish and maintain relationships with users of information in the private and public sectors, and with the general public, to enhance the relevance of the outputs and to improve their marketing.
- Before any major design or redesign, conduct extensive and user-focused consultation so as to identify content options as well as to develop public support for the program. Since outputs are not used if they are not trusted, take a very open approach when developing or revising programs so as to build confidence in the process and products.
- In determining the extent to which a survey will meet user needs, seek a reasonable trade-off between those needs and the budget, response burden and privacy considerations. The NIS may have little discretion when data content and general methodology is legislated. However, in other cases, it is worthwhile formulating alternative methodological approaches, means and modes of data collection, frequencies, geographical detail, etc., with a view to arriving at an optimum solution.
- Review ongoing surveys at regular intervals. They need to evolve, adapt and innovate so as to keep pace with changing users demands. The survey purpose needs to be re-examined and revised periodically to ensure continuing relevance to user needs as they involve or change. Sometimes the overhaul of existing surveys may be desirable to maintain the reliability of key statistical series, especially if sources of information have changed, or the ways in which they are made available are reengineered.
- Where explicit data quality targets exist, include them in the statement of survey objectives. Targets may refer to sampling errors, coverage rates, response rates, and timeliness.

References

Blanc, M., Radermacher, W. and Körner, T. (2001). *Quality and users*. Session 15.1 of the International Conference on Quality in Official Statistics, Stockholm, Sweden.

Brackstone, G.J. (1993). *Data relevance: keeping pace with user needs*. *Journal of Official Statistics*, 9, 49-56.

3.2 Concepts, Variables and Classifications

Scope and purpose

Concepts are the subjects of inquiry and analysis of interest to users. They refer to general characteristics or attributes of statistical units. *Variables* are indicators used to measure the concepts. Definitions of concepts and the corresponding variables must be unambiguous, precisely specified, and relevant to the primary analytical purposes for which the data are being collected. All categorical statistical data have to be classified for presentation and analysis and the corresponding *classifications scheme* must also be suitable for these analytical purposes.

Principles

In order to draw conclusions from a set of data, the users must have input to and knowledge of the concepts underlying the data. Use of harmonized definitions of concepts, variables and classifications assists users in comparing and integrating data. However, such definitions may need modification to meet specific user needs. A variable chosen at a point in time may become obsolete if new factors come into play.

The definitions of the concepts, variables and classifications should be documented, and any differences from standards, or from definitions used for related data, should be highlighted. This documentation is especially important for users who wish to use the data for other than purpose originally envisaged.

There is often more than one way to measure a concept. Factors to be taken into account in defining indicators include the ease of obtaining data from respondents, the respondent burden imposed, the context in which the question(s) must be asked, the collection method, the editing, imputation and weighting techniques, whether the information can be obtained from administrative records, and the costs associated with collection and processing.

Guidelines

- Specify concepts, variables and classifications clearly and justify them in terms of their intended use.
- Wherever possible, make use of the standard definitions of concepts, variables, classifications, statistical units and populations.
- In choosing naming conventions, take existing standards and usage into account. Use titles from existing standards only as defined in the standards.
- To maximize flexibility of use, code micro-data and maintain files at the lowest possible level of each classification. Aggregation to a higher level required for particular analytical purposes or to satisfy confidentiality or data reliability constraints can be made at a later stage.
- Wherever possible, use a common collapsing strategy for aggregations and define them in terms of higher level aggregations of the classification. Document differences between the standard and actual levels of classifications/aggregations used. Aim to use classifications that include detailed and the collapsed levels. Make clear to users how these relate into higher level (i.e., less detailed) classifications.

- Use standard definitions to make it possible to compare data collected from different sources and to integrate data across sources. The NIS has many standard classifications including of industries, products, occupations and geography (see references below) as well as its Glossary of Statistical Terms. In addition there are international standard classifications produced by the United Nations Statistical Office, the International Labour Office, Eurostat, and other international and regional agencies. In the absence of an official standard, examine the concepts, variables and classifications being used by related statistical programs and consult with the experts in standards when necessary.
- The NIS has produced official concordances to a number of international standard classifications. When there is a requirement to provide data to international agencies, use official concordances when they are available.
- Use standard units of observation to facilitate the comparison of data. Classifications are usually designed with particular units of observation in mind.
- Take account of derived statistical activities, such as the System of National Accounts, whose requirements in terms of variables and classifications might have a significant effect the survey.

References

Classification of Activities in National Economy - CAEN REV.1.

Classification of Activities in National Economy - CAEN REV.2.

Classification of the Products and Services Related to the Activities in the National Economy – CPSA.

Nomenclature of Industrial Services and Products – PRODRUM.

Register of the Administrative and Territorial Units of Romania – SIRUTA.

Nomenclature of Territorial Units for Statistics – NUTS.

Classification of Occupations in Romania – COR 2009.

International Standard Classification of Education – ISCED.

Emission Inventory Guidebook – CORINAIR.

Selected Nomenclature for Sources of Air Pollution – SNAP.

Standard Goods Classification for Transport Statistics - NST 2007.

List of Dangerous Goods (according to) ADR and ADN.

The Nomenclature of Products and Services – COICOP.

The Combined Nomenclature of Goods – NC.

System of National Accounts, 1993

3.3 Coverage and Frames

Scope and purpose

The *target population* is the set of units about which information is wanted and estimates are required. Practical considerations may dictate that some units be deliberately excluded, for example, small informal businesses that are impossible to access without incurring excessive cost. This gives rise to the concept of the *survey population*, the set of units obtained by excluding some units that are hard or expensive to access. Differences between the target population and the survey population are the result of deliberate restrictions on coverage. If

the two differ, valid statistical inference under probability sampling can be made about the survey population, not about the target population itself.

A *frame* is any list, material or device that delimits, identifies, and allows access to the elements of the survey population. Frames are generally of two types: area frames and list frames. Area frames are usually made up of a hierarchy of geographical units, that is, the frame units at one level can be subdivided to form the units at the next level. All the elements included in the frame constitute the *frame population*. The discrepancies between the survey population and the frame population are referred to as *coverage errors*.

Principles

The survey frame should conform to the survey population and contain minimal under-coverage and over-coverage. Frame information must be kept up-to-date. Coverage errors occur due to omissions, erroneous inclusions, duplications and/or misclassifications of the units in the survey frame.

Characteristics of the frame units (e.g., identification, contact, classification, address, size, maps in case of geographical units) should be of high quality in view of their use in stratification, sample selection, collection, follow-up, data processing, imputation, estimation, record linkage, quality assessment and analysis. Frame imperfections such as coverage errors and out-of-date characteristics are likely to bias or diminish the reliability of the survey estimates and to increase data collection costs.

Guidelines

- Test possible frames at the planning stage of a survey for their suitability and quality.
- When several frames exist, some of which are incomplete but less expensive to use and others more complete but prohibitively expensive, consider the use of multiple frames.
- Consider use of random digit dialling may be used for some telephone surveys or in combination with other area or list frames.
- For business and institutional surveys, use the Business Register (REGIS).
- For agricultural surveys, use statistical register of agricultural exploitations (REXA).
- For household surveys, use master sample for households surveys (EMZOT).
- To avoid inconsistencies, to facilitate combining estimates from the surveys and to reduce costs of frame maintenance and evaluation use the same frame for surveys with the same target population as far as possible.
- Retain and store information about sampling, rotation and data collection so that coordination between surveys can be achieved and respondent relations and response burden can be better managed. For example, record for which units' data are being collected by each survey.
- Incorporate procedures to eliminate duplication and to update for births, deaths, out-of-scope units and changes in characteristics.
- Monitor the frame quality by periodically assessing its coverage and the quality of the information on the characteristics of the units. Many techniques exist for this purpose:
 - matching the frame or a sample of the frame with comparable alternative sources, often provided by administrative records, for the survey population or subsets of it;
 - analyzing survey returns for duplicates, deaths, out-of-scope units, and changes in characteristics;
 - using specific questions on the questionnaire to aid in monitoring coverage and classification information; verifying with local authorities (e.g., regional offices, field survey staff, the survey units themselves);

- verifying the frame or subsets of it in the field (which could include verification of out-of-scope units);
- comparing the frame with a sample of units from a corresponding area frame;
- updating the frame to determine changes to it;
- checking the consistency of counts with other sources or with data from specially designed replicates;
- using evaluative information obtained from other surveys with the same frame;
- Monitor the frame between the time of sample selection and the survey reference period.
- Incorporate frame updates in the timeliest manner possible.
- Minimize frame errors through effective training of staff, an emphasis on the importance of coverage, and the implementation of quality assurance procedures of frame related activities.
- For area frames, implement map checks to ensure clear and non-overlapping delineation of the geographic areas used in the sampling design (e.g., through field checks or the use of other map sources).
- For statistical activities from administrative sources or for derived statistical activities, where coverage changes may be outside the control of the immediate manager, determine and monitor coverage, and negotiate required changes with the source manager.
- Make adjustments to the data or use supplementary data from other sources to offset coverage error of the frame.
- Include descriptions of the target and survey populations, any differences between the target population and the survey population, as well as the description of the frame and its coverage errors in the survey documentation.

References

Archer, D. (1995). *Maintenance of business registers*. In Business Survey Methods, B.G. Cox et al (eds.), Wiley-Interscience, New York, 85-100.

Colledge, M.J. (1995). *Frames and business registers: an overview*. In Business Survey Methods, B.G. Cox et al (eds.), Wiley-Interscience, New York, 21-47.

Eurostat (2003) *Business Register Recommendations Manual* (http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-BG-03-001/EN/KS-BG-03-001-EN.PDF)

Kott, P.S. and Vogel, F.A. (1995). *Multiple-frame business surveys*. In Business Survey Methods, B.G. Cox et al (eds.), Wiley-Interscience, New York, 185-203.

3.4 Sampling

Scope and purpose

Sampling is the selection of a set of units from a survey population, referred to as the *sample*, with the aim of making inferences about the population. The choice of sampling method has a considerable influence on quality, in particular sampling error. It depends many factors including the desired level of precision, detail of data output, availability of appropriate sampling frame and auxiliary variables for stratification and sample selection, estimation methods that will be used, and the available budgets.

Principles

Probability sampling means selecting units from the frame in such a way that all units have known and positive *inclusion probabilities*. Sample size is determined by the required precision

and available budget. The probability distribution that governs the sample selection, along with the stages and units of sampling, the stratification, the sample size determination and allocation are collectively referred to as the *sample design*. A combination of sampling design and estimation method (see Section 3.10) is chosen so that the resulting estimates have maximum precision for a given budget, or incur minimum cost for given precision.

Stratification consists of dividing the population into subsets (*strata*) within each of which an independent sample is selected. The choice of strata is based on the objectives of the survey, the frequency distributions of the main variable(s) of interest and the desired precision.

Sometimes the information needed to stratify the population is not available on the frame. In such cases, a *two-phase (double) sampling* scheme may be used, whereby a large sample is selected in the first phase and the required stratification data obtained. The first sample is then stratified as required and the second phase sample is selected within the first phase sample.

In case of area frames, it is not efficient to select directly the *reporting* units i.e., the units about which data will be reported. In such cases, a *two stage sampling* scheme may be used. In the first stage *clusters* (called *primary sampling units*) of potential reporting units are selected; in the second stage samples of reporting units are selected from within each of the selected primary sampling units. Budgetary or other considerations may necessitate more than two stages.

Data collected for the reporting units may, where appropriate, be supplemented at the estimation stage, with *auxiliary information* from other sources than the survey itself, such as administrative records or census projections to improve the precision of the estimates. The choice of sampling design should take into account the availability of such auxiliary information.

Guidelines

- For statistical efficiency strata should contain units that are as homogeneous as possible with respect to the information requested in the survey.
- For highly skewed populations, which means for most business populations, create a stratum of large units to be included in the survey with certainty. These large units normally account for a significant part of the estimates of the population totals.
- Most surveys produce estimates for various *domains* of interest, for example, regions. This should be taken into account in the design by stratifying according to these domains whenever possible, otherwise, it is necessary to post stratify at the estimation stage.
- For longitudinal surveys, choose stratification variables that correspond to characteristics that are stable through time.
- Where not all the data required for good stratification are available from the frame, consider the possibility of a two phase sample design. Calculate the cost of sampling at each phase, the availability of data required at each phase, and the gain in precision obtained by stratifying the first phase sample.
- For surveys based on an area frame, determine how many stages of sampling are needed and which sampling units are appropriate at each stage. For each possible type of reporting unit, consider the availability of obtaining or creating a suitable frame of such units, the ease of contact and of data collection, the quality of data, and the cost of collection.
- Conduct studies to evaluate alternative sampling methods, stratification options and allocation possibilities. The usefulness of these studies depends on the availability and age of data used for the studies and their relationship to the key survey indicators. Such data may be drawn from previous censuses, surveys or administrative data.
- When determining sample size, take into account the required levels of precision for each indicator, the sample design, the auxiliary information, the budget, and non-sampling factors including likely non-response, out-of-scope units, and attrition (in repeating and longitudinal surveys).

- Most surveys produce estimates for many indicators, and a sample that is optimised for one, may be far from optimised for others. Handle this problem by first identifying the key indicators and optimise with respect to one or two from this subset. This often requires a compromise between optimum strategies for the individual indicators.
- In determining sample allocation and size for stratified samples, take account of expected frame deficiencies, such as dead and misclassified units.
- At the implementation stage, compare the size and characteristics of the actual sample to what was expected. Compare the precision of the estimates to the planned objectives.

Repeating Surveys

- As the survey population of a repeating survey is likely to increase gradually, it is prudent to use a sample design that holds the sample size (and hence collection costs) stable. This is possible as the precision of survey estimates is more dependent on the total sample size than on the overall sampling fraction.
- The design should be as flexible as possible to deal with future changes, such as increases or decreases in sample size, restratification, resampling and updating of selection probabilities. If estimates are required for specified domains of interest (e.g., sub-provincial estimates) form strata by combining smaller stable domains (e.g., districts) so that future changes in the required domains are easier to accommodate.
- If efficient estimates of change are required and/or if response burden is a concern, use a *rotation* sampling scheme that replaces part of the sample in each period. The choice of the rotation rate is a compromise between the precision of the change estimates and respondent burden. A low rotation rate gives good precision but may lower response rates over time. It also has the additional benefit of reducing costs if the first contact is substantially more expensive than subsequent contacts.
- Develop procedures to monitor the quality of the sample design over time. Set up an update strategy for selective redesign of strata that have suffered serious deterioration.
- For longitudinal panel surveys, determine the length of time panels are in sample by balancing duration against attrition and conditioning effects. Use a design with overlapping panels (i.e., with overlapping time span) when there is a need to produce cross-sectional estimates along with the longitudinal ones.

Software

- Use generic software instead of tailor-made systems wherever possible to reduce development costs and possible programming errors. For example MICROSTRATE developed by Eurostat enables control sample overlap.

References

Cochran, W.G. (1977). *Sampling Techniques*. Wiley, New York.

Eurostat (2008) *Eurostat Sampling Reference Guidelines – Introduction to Sample Design and Estimation Techniques*. (http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-003/EN/KS-RA-08-003-EN.PDF)

Eurostat (2006) *Benchmarking through Calibration of Weights for Microdata* (http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/KS-DT-05-007/EN/KS-DT-05-007-EN.PDF)

Hidiroglou, M.A., (1994). *Sampling and estimation for establishment surveys: stumbling blocks and progress*. Proceedings of the Section on Survey Research Methods, American Statistical Association, 153- 162.

Hidiroglou, M.A, and Srinath, K.P. (1993). *Problems associated with designing sub-annual business surveys*. Journal of Economic Statistics, 11, 397-405.

Kish, L. (1965). *Survey Sampling*. Wiley, New York.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

3.5 Questionnaire Design

Scope and purpose

A *questionnaire* is a set of questions designed to collect information from a respondent. It may be interviewer-administered or respondent-completed, with data collection using paper and pencil, telephone, computer-assisted or Internet. Questionnaires have a major impact on respondent behaviour, interviewer performance, collection costs and respondent relations.

Principles

Questionnaire design (mode, content and layout) must take into account user requirements, survey operations and the nature and characteristics of the respondent population. Good questionnaires impose low response burden and remain both respondent and interviewer friendly. They ask relevant questions, permit data to be collected efficiently, with a minimum number of errors, while facilitating data coding and capture, and minimizing the amount of editing and imputation that is subsequently required.

All new and modified questionnaires must be thoroughly tested before implementation. Questionnaires in ongoing surveys should be evaluated periodically.

Guidelines

- Use words and concepts in questionnaires that are understandable to respondents. In the case of businesses, choose questions, time reference periods, and response categories that are compatible with company record-keeping practices. To the extent possible, harmonise concepts and wording with those in common use. When appropriate, reuse questions from other surveys.
- Choose question design and wording that encourage respondents to complete the questionnaire accurately. To this end, the questionnaire should not deviate from the topic of the survey, be as brief as possible, flow smoothly from one question to the next, facilitate respondents' recall and indicate appropriate information sources.
- In the introduction to the questionnaire, indicate the subject of the survey, identify the sponsor, explain the purpose, and request the respondent's cooperation. Also indicate the authority under which the survey is being conducted, the confidentiality protection measures, and any record linkage or data sharing arrangements that are envisaged. The opening questions should be applicable to all respondents, establish that the respondent is a member of the survey population, and be easy to complete.
- Design self-completed questionnaires to be attractive and easy to complete. To this end, give a positive first impression in the cover letter and front cover, and make the questionnaire appear professional and businesslike.
- If the questionnaire is to be interviewer-administered, make it interviewer-friendly.
- To minimize the possibility of reporting errors, ensure that the instructions to respondents and/or interviewers are short, clear, and easy to find. Provide definitions at the beginning of the questionnaire or in specific questions, as required (not in a separate instruction sheet). Ensure that time reference periods and units of response are clear. Use boldface print to emphasize important items. Specify "includes" or "excludes" in the questions themselves, and ensure that response categories are mutually exclusive and exhaustive.
- With respect to the questionnaire layout, provide headings for each section and instructions and answer spaces that facilitate accurate answering. Use colour, shading, illustrations and symbols to attract attention and guide respondents (or interviewers) through the

questionnaire. At the end of the questionnaire, provide space for additional comments by respondents, consider asking for the time taken to complete the questionnaire, and include an expression of appreciation to the respondent.

Testing

- The most appropriate methods for testing and evaluating a questionnaire depend upon the type and size of the survey, the content, the extent of utilisation of previous survey questions, whether the survey is to be repeated, the method of data collection, and the development schedule, budget, and availability of resources.
- Consider two or more phases of questionnaire testing, for example, testing the questionnaire at an early stage in its development, and then retesting it after making revisions based on the initial test findings. The process may be repeated several times with different methods each time.
- Use focus groups, in-depth interviews, think-aloud interviews, and behaviour coding to provide insight into how respondents may react to the questionnaire. Focus groups and one-on-one, in-depth interviews are used to test and evaluate question wording, sequencing and format. Cognitive methods are used to examine respondents' thought processes as they answer the survey questions and to ascertain whether or not they understand the questions and are able to provide accurate responses. Behaviour coding provides a systematic and objective means of examining the effectiveness of the questionnaire by analyzing the interviewer-respondent interaction. These qualitative testing methods may also help in determining questionnaire content through evaluation and exploration of key concepts.
- Conduct informal testing (pretesting) of the questionnaire to help identify poor question wording or ordering and errors in questionnaire layout or instructions, and problems likely to arise from respondents' inability or unwillingness to answer questions. Use informal testing to suggest additional response categories that can be pre-coded, and to provide a preliminary indication of the interview length.
- Hold debriefing sessions with interviewers after testing the questionnaire at which they discuss their experiences in interviewing respondents and impressions of how the questionnaire performed. Interviewers can identify potential sources of response and non-response errors as well as areas where the questionnaire can be improved.
- Use split sample testing when there is a need to determine the best of two or more alternative versions of the questionnaire. This involves an experimental design that is incorporated into the data collection process. It may be used to investigate issues such as question wording, question sequencing, and data collection procedures.
- Conduct pilot testing to observe how all the survey operations, including the administration of the questionnaire, work together in practice. The pilot test duplicates the final survey design on a small scale from beginning to end, including data processing and analysis. It provides an opportunity to fine-tune the questionnaire before its use in production.

References

- Fowler, F.J. Jr. (1995). *Improving Survey Questions: Design and Evaluation*. Applied Social Research Methods Series, 38, Sage Publications, Thousand Oaks, California.
- Gower, A.R. (1994). Questionnaire design for business surveys, *Survey Methodology*, 20, 125-136.

3.6 Data Collection and Capture

Scope and purpose

Data collection is the process of acquiring data from individuals or organisations using the questionnaire. If no data are obtained initially, or if the data obtained are incomplete or unsuitable (as identified by preliminary editing) follow-up may be required to help complete data collection (as further discussed in the next section).

Data Capture is the process that of converting data provided by respondents into electronic format suitable for subsequent processing. In the case of CAPI (Computer Assisted Personal Interviewing), CATI (Computer Assisted Telephone Interviewing) and EDR (Electronic Data Reporting), data are captured as part of the actual collection process. Otherwise data are captured in a separate operation by manual key entry or automated means such as Intelligent Character Recognition (ICR). Often this data conversion involves either manual or automated coding. Sometimes it includes transmitting the data to another location.

Principles

Data providers, especially individuals and organisations who complete questionnaires, (usually) without payment, are the NIS most valuable resource. To ensure continuing cooperation, it is essential to minimize respondent burden.

Gaps or inconsistencies in the data are best corrected by consulting respondents themselves during data collection or very soon afterwards.

Given data collection and capture operations' high impact on data quality, use of appropriate quality and performance measurement tools to manage these processes and provide objective measures to supervisors and clients is essential.

Throughout the process, appropriate steps must be taken to preserve the confidentiality and security of the information collected (as further discussed in Section 3.13).

Guidelines

- Interviewers and data capture operators are critical to the success of most data collection and capture operations. Ensure that they have appropriate training and tools.
- Exploit advances in communications and computing technology to improve the efficiency and effectiveness of data collection and captures and to reduce the costs and risks. Consider computer-assisted survey interviewing (CAPI and CATI) and electronic data reporting (EDR) via the Internet, automated data entry (using ICR) and automated coding by text recognition (ACTR).
- Establish appropriate sample control procedures for all data collection operations. Institute effective control systems to ensure the security of data capture, transmission and handling. Track the status of sampled units from sample selection through to clean edited data so that data collection managers and interviewers can assess progress at any point in time.
- Control paper questionnaire delivery operations in mail surveys to ensure that each unit that has been selected receives the appropriate questionnaire. Once the questionnaire is returned, verify the accuracy of the coverage information and the quality of the data provided. Follow-up interviews may be needed in some cases. When no questionnaire is received, follow-up activities are necessary to establish the status of the unit (e.g., active or inactive) and to obtain the missing information.
- When collecting data, ensure that the appropriate person within the responding household or organisation is contacted at the appropriate time so that the information is readily available. Allow the respondent to provide the data using a method and format that is convenient to them or their organisation. This helps improve response rates and data quality.

- Ensure that procedures are applied to all units consistently. Automation is desirable and releases staff to work on difficult cases. Centralize processing to reduce costs and make it easier to take advantage of a pool of expert knowledge. Given that changes to procedures may be required in the future, design processes that can be easily modified.
- Monitor the frequency of edit failures and corrections by stratum, collection mode, processing type, and data item. This will help in evaluating the quality of the data and the efficiency of the editing function.
- Expenditure, performance and quality measures gathered during the data collection operation enable the survey manager to make decisions regarding the need for modification or redesign of the process. Track actual costs of postage, telephone calls, questionnaire production, computing, and person-days. Important quality measures include response rates, processing error rates, follow-up rates, and counts of non-response by reason.
- Manual data capture from paper questionnaires is subject to keying errors. Incorporate on-line edits so that the data capture operator can correct keying errors and record them for later review and analysis.
- Use statistical quality control methods to assess and improve the quality of clerical collection and capture operations, in particular data capture and coding. Collect and analyze quality control results to identify the major causes of error and report to managers, staff, subject-matter specialists and methodologists.
- Use subsequent survey processes to gather useful information regarding quality that can serve as signals that collection and capture procedures and tools may require changes for future survey cycles. For example, editing and evaluation (see Sections 3.8 and 3.12) may suggest collection related problems.

References

Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L. and O'Reilly, J. (eds.) (1998). *Computer Assisted Survey Information Collection*. Wiley, New York.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. Wiley, New York.

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. and Trewin, D. (eds.) (1997). *Survey Measurement and Process Quality*, Wiley, New York.

Mudryk, W. and Xie, H. (2002). *Quality control application in ICR data capture for the 2001 Census of Agriculture*. Proceedings of the Section on Quality and Productivity, American Statistical Association, 2424-2429.

Mudryk, W., Burgess, M.J. and Xiao, P. (1996). *Quality control of CATI operations at Statistics Canada*. Proceedings of the Section on Survey Research Methods, American Statistical Association, 150-159.

3.7 Non-Response

Scope and purpose

Despite the best efforts of survey staff to maximize response, some non-response will occur. Non-response can be of two types. *Item non-response* is where a particular data item on a questionnaire has not been completed, and *unit non-response* is where no response has been received from the unit or the response is so limited as to be effectively non-existent.

The two approaches for dealing with missing data are imputation (Section 3.9) and sampling weight adjustment (Section 3.10). Typically, weights are adjusted for unit non-response and imputation used for item non-response.

Non-response has two effects on data. First, it introduces bias in estimates when non-respondents differ from respondents and correction methods are the same. Second, it increases sampling variance because the effective sample size is reduced.

Principles

An effective respondent relations program and a well-designed questionnaire are critical elements in maximizing response.

Non-response is monitored and results fed back survey staff for immediate and future actions and reported to users.

The effort made to get a response in any given case depends upon budget, timeliness considerations, and likely impact on quality of non-response bias. The appropriate amount of research devoted to development of non-response adjustment techniques is likewise influenced by consideration of budget, time, use of data, and risk of bias.

Guidelines

- High response rates require an integrated approach, along the following lines:
 - well defined frame units;
 - up to date survey frame in terms of coverage and contact data;
 - well defined reference period and appropriate timing and length of collection period;
 - appropriate method(s) of data collection; minimal response burden (in terms of length of questionnaire, difficulty of questions, frequency of collection); attention to sensitivity of subject matter (sensitivity of subjects);
 - effective follow-up methodology and anticipation of difficulties in tracing respondents who have moved – use of pre-test and/or results from similar surveys to establish expected response rates;
 - effective use of respondent incentives;
 - prior experience with same type of survey, and demonstrated ability of collection staff;
 - reasonable workloads for collection staff;
 - well established knowledge of and relationships with respondents;
 - sufficient budget and its appropriate allocation to the various operations.
- Follow up non-respondents to the extent the budget permits. Follow-up increases response rate and helps determine whether respondents and non-respondents are similar with respect to characteristics measured.
- Follow-up is particularly important for longitudinal surveys where the sample is subject to accumulating attrition due to non-response at each survey occasion. Facilitate high quality tracing using administrative data, city and telephone directories, and local knowledge. Verify/update contact information on each survey occasion and ask units to notify the NIS if changes occur between survey cycles.
- Prioritize follow-up activities. For example, in business surveys, follow-up large or influential units first. Likewise, give a higher priority to non-responding units in domains with high potential for non-response bias. Use a score function to prioritise follow-up.
- Record and monitor the reasons for non-response (e.g., refusal, non-contact, temporarily absent, technical problem). The degree of non-response bias may differ depending on the reason. Monitor non-response trends by reason.
- Try to determine differences (if any) between respondents and non-respondents by linking to external data sources (for example, administrative data files) and by examining the responses of the non-respondents who became respondents during a follow-up. For repeating surveys, known characteristics of respondents for one cycle/wave of the survey

can be used to compare characteristics of respondents and non-respondents in subsequent cycles/waves.

- Attempt to evaluate the extent to which the non-response adjustment procedures correct for the potential bias. Take non-response into account when producing estimates and their associated variance estimates.
- Record unweighted and weighted non-response rates at the estimation stage.
- Inform users of non-response rate when providing estimates.

References

Cialdini, R., Couper, M. and Groves, R.M. (1992). *Understanding the decision to participate in a survey*. Public Opinion Quarterly, 56, 475-495.

Federal Committee on Statistical Methodology (2001). *Measuring and reporting sources of error in surveys*. Statistical Policy Working Paper 31. See also <http://www.fcsm.gov>.

Groves, R.M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (2002). *Survey Non-response*. Wiley, New York.

3.8 Editing

Scope and purpose

Data *editing* is the application of checks to detect missing, invalid or inconsistent data item entries or to point to data records that are potentially in error. Some of these checks are based on logical relationships that follow directly from the data item definitions and relationships. Others are empirical in nature – feasibility checks, outlier detection techniques. Checks may be based on data from previous cycles of the same survey or from other sources.

Editing encompasses a wide variety of activities, from interviewer field checks, computer-generated warnings during data collection or capture, identification of units for follow-up, error localization for the purposes of imputation, to data validation. The last two topics are addressed in Sections 3.9 and 3.12.

Principles

The goals of editing are threefold:

- to clean the data;
- to provide information data quality; and
- to provide the basis for future improvement of the survey questionnaire and operations.

Avoid spending a disproportionate proportion of survey resources cleaning the data.

Fatal errors (invalid or inconsistent entries) should be removed from the data to facilitate further automated data processing and analysis and to ensure the NIS's credibility. Identification of fatal errors can and should be automated. However, there is a danger of overusing *query edits* (those pointing to entries or records that may be in error), particularly given that editing may be the single most expensive survey operation, accounting for perhaps a quarter of the total budget.

Edit hit rates, defined as the proportion of edits that actually reveal true errors are often quite low, for example 20-30%. Furthermore, it is not uncommon for a few error corrections to be responsible for the majority of changes in the estimates due to editing. When the impact of most query edits on the final estimates is negligible *over-editing* is said to have occurred.

Editing can play a valuable role in providing information for subsequent improvement of the survey process - in sharpening definitions, improving the questionnaire, evaluating data quality, and identifying non-sampling error sources.

Guidelines

- Ensure that the edit rules are internally consistent.
- Reapply edits to units to which corrections have been made to ensure that no errors were introduced by the correction process itself.
- Edit selectively. Priorities may be set according to types and/or severities of errors, and/ or according to the importance of the data items and/or reporting units.
- For business surveys, put in place a strategy for selective follow-up. The use of a score function is a means for ensuring resources are systematically concentrated on the most important units, key indicators, and severe errors.
- Consider reducing editing effort and redirecting resources into activities with a higher pay-off, for example, data analysis and response error analysis.
- Assign a high priority to learning from the editing results. To reduce errors, practice error prevention rather than error correction. To this end, ensure editing takes place as early in the survey process as possible, preferably while the respondent is still available, for example, through the use of computer-assisted telephone or personal or self-interview methods.
- Edits cannot possibly detect small, systematic errors reported consistently in repeated surveys. Tighter edits are not the solution. Such errors are best detected and corrected through analysis and review of concepts and definitions, post-interview studies, data validation, and data confrontation with other data sources.
- Identify *outliers* (extreme data values) in a survey period or across survey periods. They are a warning sign of potential errors.
- When conducting follow-ups, do not overestimate the respondents' ability to report or correct earlier reports. Their memories are limited and they may provide different data.
- The editing process is often very complex. Ensure procedures are well described and up-to-date procedures, provide appropriate training to all staff involved, and monitor the work itself. Consider using formal quality control procedures.
- Automate editing to the fullest extent possible. While some manual intervention may be necessary, generalized, reusable software is particularly useful for this purpose. However, do not be tempted to increase the scope and volume of checks simply because they can be automated. Query checks that make little difference to the estimates should be excluded.

References

Bankier, M., Lachance, M. and Poirier, P. (1999), *A generic implementation of the New Imputation Methodology*. Proceedings of the Survey Research Methods Section, American Statistical Association, 548-553.

De Waal, T., Van de Pol, F. and Renssen, R. (2000). *Graphical macro editing: possibilities and pitfalls*. Proceedings of the Second International Conference on Establishment Surveys, Buffalo, New York.

Granquist, L. and Kovar, J.G. (1997). *Editing of survey data: how much is enough?* In *Survey Measurement and Process Quality*, Lyberg et al (eds.). Wiley, New York, 415-435.

Latouche, M. and Berthelot, J.-M. (1992). *Use of a score function to prioritize and limit recontacts in editing business surveys*. *Journal of Official Statistics*, 8, 389-400.

3.9 Imputation

Scope and purpose

Imputation is the process replacing values that are missing or designated invalid or inconsistent, thereby creating internally consistent records, whilst preserving as much of the original data values as possible.

Ideally, missing and invalid and inconsistent entries should be resolved at the data collection or follow-up stages. However, this is impossible in practice because of respondent burden, cost and timeliness constraints.

Some methods of imputation may distort underlying data item frequency distributions and the relationships between data items. This must be taken into account when producing estimates and associated variance estimates.

Principles

Imputation may be automated, manual or a combination of both.

Good imputation processes:

- depend on access to all the micro-data and good auxiliary information;
- ensure that imputed records are internally consistent;
- result in the minimum number of changes to data items values in a record;
- limit the bias caused by not having observed all the desired values;
- are automated, objective, reproducible and efficient;
- leave an audit trail for evaluation purposes.

Imputation methods can be classified as either deterministic or stochastic

- *Deterministic imputation* methods include logical imputation, historical imputation, mean imputation, ratio and regression imputation and single donor nearest-neighbour imputation. These methods can be further divided into methods that rely solely on deducing the imputed value from data available for the non-respondent and/or other auxiliary data about the respondent, and those that make use of the observed data for other responding units for the given survey.
- *Stochastic imputation* methods include the hot deck, nearest neighbour imputation where a random selection is made from several near neighbours, regression with random residuals, and any other deterministic method with random residuals added.

Guidelines

- Determine which auxiliary data items can best explain the non-response mechanism(s) and use them to enrich the imputation method. Study the quality and appropriateness of available data items and select which to use as matching data items and which to use to construction imputation classes.
- Take into account the type of estimates to be produced, level or change, large aggregates or small domains, cross-sectional or longitudinal data.
- It is often appropriate to use several different types of imputation methods in an ordered sequence.
- Ensure imputed records are internally consistent.
- Ensure imputed records resemble the original records as closely as possible.
- Test the imputation approach. Consult subject matter experts.

- When donor imputation is used, try to impute data for a record from as few donor records as possible. Operationally, this often means one donor record per section of the questionnaire. Also allow equally good donors an equal chance of being selected and avoid systematically selecting any one donor.
- For large surveys, it may be necessary to impute data in two or more passes, rather than in a single pass, so as to reduce computational costs or because there are many missing or incorrect fields on records.
- Flag imputed values and clearly identify the methods and sources of imputation. Retain the previous and imputed values for evaluation purposes.
- Evaluate the degree and effects of imputation, in particular how to measure sampling variance under imputation.
- Consider the degree and impact of imputation when analyzing data. Imputation may have a significant impact on data distributions. For example, values in one domain have moved systematically up, while values in another domain have moved down by an offsetting amount leaving totals looking the same. The greater the degree and impact of imputation, the more judicious analysts must be in using the data. In some cases, analyses may be better conducted excluding imputed values.

Software

- During the development of the imputation methodology make use of any generalised imputation systems that are available. At the NIS, estimation of quantitative data items is currently done using SAS macros `mainmean.sas` and `macmean.sas`, and of ordinal and nominal variables by SAS macros: `ini_execute.sas` and `execute.sas`.

References

Bankier, M., Lachance, M. and Poirier, P. (1999). *A generic implementation of the New Imputation Methodology*. Proceedings of the Survey Research Methods Section, American Statistical Association, 548-553.

Fellegi, I.P. and Holt, D. (1976). *A systematic approach to automatic edit and imputation*. Journal of the American Statistical Association, 71, 17-35.

Kovar, J.G., and Whitridge, P. (1995). *Imputation of business survey data*. In Business Survey Methods, B.G. Cox et al. (eds.), Wiley, New York, 403-423.

Lee, H., Rancourt, E. and Särndal, C.-E. (2002). *Variance estimation from survey data under single imputation*. In Survey Non-response, R.M. Groves et al. (eds.), Wiley, New York, 315-328.

3.10 Estimation

Scope and purpose

Estimation is the process generating estimates of population parameters based on sample data. Examples of parameters include simple descriptive statistics such as totals, means, ratios and percentiles, and more complicated analytical statistics such as regression coefficients. Inference may be in the form of point estimates, confidence intervals, results of hypothesis tests. Point estimates are usually accompanied by estimates of *precision*.

Where population parameters are particular functions of population totals, their estimators are generally the corresponding functions of the estimated population totals.

Auxiliary data can be used to improve the precision of the estimates.

Principles

The *total survey error* of an estimate is the amount by which it differs from the true value of population parameter being estimated. It can be viewed as the sum of the *sampling error* and the *non-sampling error*. The sampling error is the error associated with basing the estimate on data from the sample only. It includes sampling bias and sampling variance. Non-sampling errors include all other causes of lack of precision, including coverage errors, measurement errors and non-response errors, as further discussed in other sections.

The sample design and the estimation method jointly determine the properties of the sampling error. The goal is to choose an estimator (i.e., estimation method) that give minimum bias and variance and is consistent, meaning that sampling error is zero when the sample comprises the whole population.

When no useful auxiliary data are available for estimation, the most natural choice of estimator is the *Horvitz-Thompson estimator*. It assigns weights to units that are the inverses of the inclusion probabilities of the units. Such weights are called a *sampling weight*. A sampling weight can be interpreted as the number of times that a unit should be replicated to represent the full population.

An estimator can be improved when auxiliary are available. *Calibration* is a procedure by which auxiliary data are incorporated. It adjusts the sampling weights by multipliers, known as *calibration factors*, that make the estimates agree with known totals. The resulting weights are called *calibration weights* or *final estimation weights*. These calibration weights will generally result in estimates that are consistent and that have a smaller sampling variance than the Horvitz-Thompson estimator. *Bounding methods* are used to ensure that calibration does not produce negative or very large weights.

As a result of non-response, the observed sample is smaller in size than the original sample selected. To compensate for non-response *reweighting* is usually used. It consists of adjusting the sampling weights by *non-response adjustment factors* before calibration. The basic principle in computing the non-response adjustment factors is to use the inverse of the response probabilities. However, these are unknown and must be estimated. The key to reducing non-response bias is to obtain a useful non-response model by taking full advantage of auxiliary data.

Outliers may lead to unreliable estimates for continuous variables. Outliers might be due either to extreme values of some data items, or to very large weights, or both. Some form of outlier detection and adjustment is required.

For longitudinal surveys, two sets of estimation weights are usually derived: the *longitudinal weights* and the *cross-sectional weights*. The longitudinal weights refer to the population at the initial selection of the longitudinal sample. These weights are usually adjusted to take into account the attrition of the sample over time. They are used when performing analysis of the longitudinal data. The cross-sectional weights are related to the population established at each survey wave. These weights are normally used to produce point estimates, or differences of point estimates between two time periods. Because of the changes in the population over time, the two sets of weight are generally different.

In repeated or longitudinal surveys with a large sample overlap between occasions, consider the use of estimation methods that exploit the correlation over time. In essence these methods treat the data from previous occasions as auxiliary variables. One such method is referred to as *composite estimation*.

Guidelines

- Incorporate sampling weights in the estimation process. This implies reflecting all relevant aspects of the sample design - stratification, clustering, and multi-phase or multi-stage information - in the estimation procedures
- Whenever auxiliary data are available for sample units, together with known population totals for such data, consider using calibration so that the weighted auxiliary data add up to these known totals. This may result in improved precision and lead to greater consistency between estimates from various sources.
- When the original classification of sampling units has changed between the time of sample selection and estimation, consider *domain estimation* so that the new classification is reflected in the estimates. Domain estimation refers to estimation for specified subsets of the population (or *domains*) of interest. Often the units falling in these subsets have not been, or could not have been, identified before sampling. Estimation in the presence of dead or out-of-scope units in the sample is an example of domain estimation.
- The quality of non-response adjustment factors depends on assumptions made in the (explicit or implicit) non-response model. The model should be validated making use of auxiliary data correlated with the propensity to respond. To obtain some protection against model failure, form non-response adjustment classes and estimate the response probabilities by the response rates within these classes.
- Two-phase sampling theory can be used to estimate the variance for various estimators incorporating the non-response adjustments.
- When appropriate, use two phase sampling to improve estimation by incorporating auxiliary data. Auxiliary data can be used (a) to stratify the second-phase sample, (b) to improve the estimate using a difference, ratio or regression estimator, or (c) to draw a sub-sample of non-responding units. A fairly general approach to two-phase sampling is when auxiliary data are incorporated in the estimation process via the Generalized Regression Estimator.
- The requirements for small domains, such as small areas, are best incorporated at the sample design stage. However, if this is not possible at the design stage, or if the domains are only specified later, consider special estimation methods at the estimation stage. These methods borrow strength from related domains to minimize the mean square error of the resulting estimator. Examples are small area estimators.
- To deal with outliers use outlier-resistant (robust) estimators.
- Whenever possible, use generalized estimation software in preference to tailor-made systems. Possible software packages include SAS, FOX, SUDAAN, PC CARP, STATA.

References

Chambers, R.L., Kocic P., Smith P. and Crudas M. (2000). *Winsorization for identifying and treating outliers in business surveys*. Proceedings of the Second International Conference on Establishment Surveys, June 17-21, 2000, Buffalo, New York, 717-726.

Cochran, W.G. (1977). *Sampling Techniques*. Wiley, New York.

Deville, J.-C. and Särndal, C.E. (1992). *Calibration estimators in survey sampling*. Journal of the American Statistical Association, 87, 376-382.

Fuller, W.A. (2002). *Regression estimation for survey samples*. Survey Methodology, 28, 5-23.

Hidiroglou, M.A. and Särndal, C.E. (1998). *Use of auxiliary information for two-phase sampling*. Survey Methodology, 24, 11-20.

Lee, H. (1995). *Outliers in business surveys*. In Business Survey Methods, B.G. Cox et al. (eds). Wiley, New York, 503-526.

Rao, J.N.K. (1999). *Some recent advances in model-based small area estimation*. Survey Methodology, 25, 175-186.

Singh, A.C., Kennedy, B. and Wu, S. (2001). *Regression composite estimation for the Canadian Labour Force Survey with a rotating design*. Survey Methodology, 23, 33-44.

Singh, M.P., Hidioglou, M.A., Gambino, J. and Kovacevic, M. (2001). *Estimation methods and related systems at Statistics Canada*. International Statistical Review, 69, 461-486.

Thompson, M.E. (1997). *Theory of Sample Surveys*. Chapman and Hall.

3.11 Seasonal Adjustment and Trend-Cycle Estimation

Scope and purpose

Sub-annual macro-economic series are key for economic policy-making. However, they are influenced by seasonal fluctuations and calendar/ trading-day effects that mask relevant short and long-term movements and impede a clear understanding of underlying economic phenomena. Main aim of seasonal adjustment is to filter out usual seasonal fluctuations and typical calendar effects.

A time series is viewed as having four components – a trend, a cyclical effect, a seasonal effect, and an irregular component.

- The *trend* is the underlying long-term movement lasting many years.
- The *cycle*, usually called the business cycle, is a quasi-periodic oscillation lasting in the order of a few years.
- The trend and cycle are usually estimated in combination.
- The *irregular component* represents those random variations that are unforeseeable movements related to events of all kinds.

Seasonal adjustment consists of estimating seasonal factors and applying them to a time series to remove the *seasonal variations*. These variations represent the composite effect of climatic and institutional factors that repeat with a certain regularity within the year.

A seasonally adjusted series consists of the *trend-cycle* and the irregular component. Smoothing the seasonally adjusted series to eliminate the irregular component produces the *trend-cycle estimates*, which provide a reading of the underlying business cycle - the direction of movement, the amplitude of the cycle and, especially, the timing of turning points.

In the context of seasonal adjustment, the aim of *pre-treatment* is to ensure reliable estimation of seasonal component by detecting and correcting the series for specific known problems prior to estimation. Pre-treatment includes graphical analysis, calendar adjustments (for trading days, moving holiday and EU calendar), outlier detection and correction, treatment of missing observations, and selection of *model and decomposition scheme*.

Principles

A time series is seasonally adjusted only when there is evidence that the series is influenced by seasonal forces and the series contains identifiable seasonality. *Identifiable seasonality* is defined as a seasonal pattern that is not obscured by a high degree of irregular fluctuations and thus can be identified reliably.

Ideally the seasonal adjustment procedure does not leave any residual seasonality in the series and the resulting seasonally adjusted series is somewhat smoother than the original. The trend-cycle series is much smoother as the irregular component has been eliminated.

The last few points in the seasonally adjusted series are the ones most prone to error and hence to revision as more points become available. Ideally, revisions to the seasonally adjusted estimates should be minimal.

For aggregate series resulting from a combination of component series, those component series that contain identifiable seasonality should be seasonally adjusted, the remainder should be left unadjusted. The aggregate series itself should be adjusted so that it does not contain residual seasonality and it is relatively smooth. It may be adjusted by indirect or direct method.

- In the *indirect method*, the seasonally adjusted components and the unadjusted ones are combined to obtain the seasonally adjusted aggregate.
- In the *direct method*, the aggregate is seasonally adjusted, and its additive property (of being the sum of the components) is restored by raking/calibration of the seasonally adjusted components.

ESS Guidelines on Seasonal Adjustment

The ESS Guidelines published in 2008 provide definitive and comprehensive description of all aspects of seasonal adjustment. They focus on two approaches:

- TRAMO-SEATS - supported by Banco de España
- X-12-ARIMA - supported by US Bureau of Census

They cover pre-treatment of series, revision policies, quality, and documentation. They discuss the limitations of seasonal adjustment, e.g. short time series and they provide three sets of alternatives: A - best approach; B acceptable approach; C practices to be avoided

Guidelines

- Make use of the recently published ESS Guidelines on Seasonal Adjustment.
- The choice between the two recommended methods of seasonal adjustment is a matter of experience, subjectivity and characteristics of the time series.
- Before seasonally adjusting a time series for the first time, conduct a thorough analysis to assess if seasonality is identifiable. Publish the series in seasonally adjusted form only if it is identifiable. Check the situation subsequently every few years,
- The frequency of revisions should be minimized. Although revisions arise with each new data point added, implement revisions only when they bring about improvement in the estimates, that is, when the revised estimate moves appreciably closer to the final estimate.
- Wherever seasonally adjusted figures pertaining to the same economic activity are published, coordinate the seasonal adjustment options applied by the areas involved, and make every effort to treat related series in a consistent manner.
- On an annual basis, revise the seasonally adjusted values for the last three years when data for the first period (month, quarter) of the next year becomes available. When seasonally adjusted values are derived with year-ahead (forecast) seasonal factors, the annual revision should apply to the last four years.
- For aggregate series comprising combinations of component series choose the direct approach but only if the calibration/raking does not distort too much the month-to-month (quarter-to-quarter) movements of the component series.
- Wherever seasonally adjusted figures pertaining to the same economic activity are published, coordinate the seasonal adjustment options applied by the areas involved. For example, when possible, make consistent choices between direct and indirect adjustment of composite series, and ensure that extreme occurrences in the time series brought about by events such as plant closures, strikes, natural disasters, etc. are treated in a consistent fashion by the different areas.

- For the seasonal adjustment of recent observations, use a *concurrent seasonal factor*. This is a factor obtained using all the available data points. An exception to this guideline may apply when the most recent observations have been subjected to historically large revisions. In this case *year-ahead (forecast) seasonal factors* based on data that ended at the end of the previous year may be more appropriate.

When Using X-12-ARIMA Trend-Cycle Estimation

- Use the daily weights that are automatically estimated for series with trading-day variations. During the current year, keep them fixed by supplying them as prior daily weights. They will be modified at the next annual revision. Exceptions to this guideline may occur when a priori daily weights can be provided by subject matter experts based on better knowledge of the series in question
- Use the Easter effect factors calculated automatically by the program.
- Use the automatic ARIMA extrapolation subroutine whenever possible. If none of the built-in models is selected, an ARIMA model has to be provided.
- Use the Henderson moving averages to produce the trend-cycle estimates. To ensure that the trend-line lies within the scatter plot of the seasonally adjusted series, apply the Henderson moving averages to the published seasonally adjusted series.
- Before applying the trend-cycle estimator, extend the seasonally adjusted series with one year of forecasted values from an ARIMA model fitted to the seasonally adjusted series.
- Apply the Henderson moving averages to the extended seasonally adjusted series from which the extremes have been previously corrected.
- Use the Henderson moving average automatically selected by the X12-ARIMA program: based on the value of the irregular/trend-cycle ratio.
- Ensure the trend-cycle estimates are consistent with published seasonally adjusted estimates.
- Inform users that the last few trend-cycle estimates (especially the very last estimate) are subject to large revisions. There may even be a reversal of movement when one more data point is added to the series. The uncertainty associated with the estimates around the end can be indicated, for example, by a dashed line on the trend graph or by a written caveat to users.

References

ESS Guidelines on Seasonal Adjustment (2009).

3.12 Data Quality Evaluation

Scope and purpose

Data quality evaluation is the process of determining whether final outputs meet the original objectives of the survey/ statistical process. Evaluation results enable users to better interpret survey results and the NIS to improve quality of the current and/or subsequent survey cycles.

There are two broad methods of evaluating data quality:

- *Certification* or *validation* is the process whereby data are analysed before official release with a view to avoiding gross errors and eliminating obviously poor quality data. This process frequently coincides with an interpretative analysis of the data and is invariably subject to tight time constraints, and thus can involve only methods that yield rapid results.

- *Sources of error studies* provide quantitative information on the specific sources of errors in the data. Typically, the results of these studies are only available after the official release of the data.

Principles

Users must be able to determine to what extent data errors are likely to affect their use of the data. However, given that they are rarely independently capable of evaluating accuracy themselves, it is incumbent on each agency to evaluate data quality and provide users with the results in a usable form and on a timely basis.

Data quality evaluations are also useful to the NIS itself. To the extent that errors can be associated with particular stages of the survey process, an evaluation can lead to quality improvement in the next survey cycle and/or in other similar surveys.

The sources of error that must be evaluated include the following

- *Coverage errors* which consist of omissions, erroneous inclusions, and duplications in the frame used to conduct the survey. They may translate into a negative or positive bias in the data, and the impact may vary depending on particular survey population. Errors in coding, notably in applying industrial and geographical classifications, may lead to coverage error.
- *Sampling errors* occur when survey results are obtained from a sample rather than the population as a whole. They may also include *estimation errors* due to estimators that, by design or otherwise, create bias, for example, some small area estimators.
- *Non-response errors* occur when there is no response to one or all of the survey questions. Non-response leads to an increase in variance as a result of a reduction in the actual size of the sample and the recourse to imputation, and it produces a bias if the non-respondents have characteristics of interest that are different from those of the respondents.
- *Measurement errors* occur when the response provided differs from the real value. Such errors may be random, resulting in increased variance, or systematic resulting in bias. They may be attributable to the respondent, the interviewer, the questionnaire, the collection method or the respondent's record-keeping system.
- *Processing errors* occur when checking, coding, entering, imputing, and tabulating data. Like measurement errors, processing errors may lead to variance and/or bias.

There are two very different kinds of processing errors. The first type concerns *micro-data*. The second type concerns *macro-data* and involves *serious mistakes in calculation or presentation of aggregates that are not found until after publication*. Mistakes are the errors most visible to the public, typically receiving a lot of negative attention. Examples are when the methodology is not applied correctly, when the wrong number is inadvertently included in a press release, and when analytical presentations or diagrams give wrong impressions. They may occur in any stage in the production of statistics: programming, calculation, report writing, editing of manuscripts, etc.

Guidelines

- Determine the extent of data quality evaluation required for the survey/ statistical process considering the following factors: data uses; risk of errors and impact of errors on data use; quality variation over time; cost of the evaluation in relation to the total cost of the program; needs for improvement in quality and/or efficiency; usefulness of results to users; and whether or not the survey will be repeated.
- In the case of census and survey data, minimum requirements include measuring or evaluating coverage errors, response or imputation rate and (if dealing with a sample survey) measurement of sampling errors for key indicators.
- Provide a quality evaluation based on expert opinion when quantitative measurements are limited by the nature of the outputs, time constraints, cost, or technical feasibility.

- Make data quality evaluation part of the overall survey design as the information needed is often best collected during the survey process. Data quality reports should be included in the dissemination schedule for the survey.
- In the case of repeated surveys, it is likely not necessary, nor even possible, to produce a detailed quality evaluation on every occasion. However, a periodic review should be scheduled at regular, say annual intervals, not just when problems arise.
- Involve users in planning the evaluation and discussing the results. When circumstances permit, involve them in the actual evaluation process.
- Among evaluation methods, consider:
 - checking coherence in relation to external data sources – for example: other surveys, other iterations of the same survey, administrative data;
 - checking internal coherence – for example: by calculating ratios that are known to be within certain limits (male-female ratios, average values of properties, etc);
 - analysing largest units individually as regards their contributions to overall estimates (generally applied to business surveys);
 - calculating data quality indicators – for example: non-response rates, imputation rates and coefficients of variation;
 - holding feedback sessions with staff involved in data collecting and processing;
 - reasonableness checks by well-informed experts, including a pre-release external review.
- The type and number of mistakes that have been officially recognised and have resulted in revisions should be analysed and presented for several years back.
- Procedures to minimise the risk of mistakes and for handling the situation when mistakes have been discovered, should be developed and documented.

References

Biemer, P., Groves, R.M., Mathiowetz, N.A., Lyberg, L. and Sudman, S. (eds.) (1991). *Measurement Errors in Surveys*. Wiley, New York.

Lessler, J.T. and Kalsbeek, W.D. (1992). *Nonsampling Errors in Surveys*. Wiley, New York.

3.13 Disclosure Control

Scope and purpose

When publishing data the NIS is obligated by law to protect the *confidentiality* of respondents' information. This means ensuring that no data item for any individual (person, business) is *disclosed*, i.e., can be precisely deduced, or inferred within a narrow range

In a table, a data cell is said to be *sensitive* if it results in disclosure, i.e., reveals data about an individual irrespective of whether or not the data item is actually considered sensitive by the respondent. (For example, for a business respondent, income may be sensitive, whereas district of operation may not.)

Disclosure control refers to measures taken to adjust output data so that confidentiality is protected. As it usually results in modifying or suppressing some data, the goal is to ensure confidentiality protection with minimum loss of output.

From confidentiality perspective there is an important distinction between

- micro-data outputs - anonymized for individual records and

- aggregate/tabular data outputs – in form of tables of frequencies (counts) or tables of magnitudes (like income).

Public perception of vigilance in protecting confidentiality is as important as reality of what is done.

Principles

The NIS has set of Rules on the Confidentiality of Statistical Data, also a policy for the Protection of Classified Information.

Disclosure Control for Tables

The rules for determining whether or not any given cell is sensitive are based on *disclosure criteria*. Two types of criteria are used. One is depends on the number of respondents contributing to the cell, the other depends the relative sizes of contributions. An example of the former is that the number of respondents contributing to a cell must exceed three. (Less than three is unacceptable since, if there are only two respondents, one of the respondents can derive the value for the other respondent by simple subtraction.)

A general example of a rule depending on relative size is the (n,k) rule in which a cell designated as sensitive if the largest n respondents account for $k\%$ or more of the total cell value. Another rule is the p -percent rule. For example, under a $p=15$ rule, a cell is designated sensitive if the sum of the contributions of the third largest and all lower ranking respondents' values is less than 15% of the largest respondent's value. This ensures that the unit with the second largest value cannot estimate the largest unit's value too closely.

Cells that are deemed sensitive according to the disclosure criteria have to be modified or suppressed. There are many possible *cell suppression* rules. Suppressing only the sensitive cells is not sufficient when marginal totals are published, because the value in the suppressed cell can be deduced. To preserve confidentiality it is necessary to undertake *complementary suppression*, which entails identifying all the other *complementary* cells that have to be suppressed in order to protect the sensitive cell. Sophisticated software exists to identify complementary cells.

There are alternatives to cell suppression. One method is to collapse categories so that fewer or none of the cells are sensitive. Another method is to add noise to the data, either in the table itself or the underlying micro-data. This may be done, for example, by rounding, though conventional rounding, i.e., to the nearest higher level unit of measure, may not add enough noise to give protection.

There are several generalized disclosure control software packages available including CONFID (Statistics Canada, 2002) and ARGUS (Hundepool et al, 2002).

In the case of multiple releases from the same database, especially in the case of ad hoc requests, it is difficult to find software that can take care of all possible confidentiality concerns and some manual intervention is necessary.

Disclosure Control for Micro-data Files

Even in the absence of obvious identifying information such as names and telephone numbers, micro-data files may contain a number of data items, termed *key data items*, which in combination can identify individuals uniquely. This means effective disclosure. Even if the individuals identified are not truly unique, or if they have been wrongly identified, the *appearance* of a disclosure can be harmful to the NIS.

The number and nature of the key data items determine the disclosure risk. Some characteristics such as detailed location, or exact income, are considered to present a high disclosure risk. Micro-data files for businesses are rarely released because of the heterogeneous nature of business data. The presence of hierarchical relations between units may also increase disclosure risk. The situation is further complicated for longitudinal surveys as the set of key variables may evolve over time.

Disclosure risk also depends upon methodology.

- It increases with sampling rate. In the case where the sampling rate is 100% for the entire population, or for identifiable strata, micro-data should not be released.
- The presence of measurement errors and of imputed values lowers disclosure risk.

There are two basic *restricted data* methods for micro-data disclosure control.

- *Data reduction methods*, which include sampling, broadening classification categories, removing selected data items for some respondents, and removing selected respondents.
- *Data modification methods* include adding random noise, swapping data items between records, replacing small groups of units with average values, and replacing data for some respondents by imputed values.

As an alternative to releasing micro-data, *restricted* access procedures may be established that permit selected research workers to access to micro-data under controlled conditions.

Guidelines

Tabular Data

- Determine the sensitivity of each cell. Do not release a data table if it contains values for cells that are considered to be sensitive.
- Determine if zero frequency cells represent a problem. Zero frequency cells may reveal sensitive information in tables of magnitude data.
- Consider alternative methods to cell suppression, for example the use of random rounding.
- In the case of multiple releases from a single database, use common sense to supplement the results obtained from confidentiality protection software.

Micro-data

- Assess the risk of disclosure for micro-data files and apply appropriate data reduction or modification methods.
- Ensure that all public use micro-data files are reviewed by a senior NIS Committee before release
- For longitudinal surveys define the release and confidentiality protection strategy in advance of the first wave. This may require making assumptions about how the key data items and how they will evolve over time.
- As an alternative to releasing micro-data files, consider using *restricted access* methods such as remote access or research data centres.
- Use generalized disclosure control software in preference to custom-built systems whenever possible.

NIS Confidentiality Rules

To prevent the dissemination of confidential statistical data, the following rules must be observed.

- Individual statistical data regarding a natural or legal person may not be disseminated unless the person expresses his/her approval in writing.
- Aggregate data for two natural or legal persons may not be disseminated unless they expressed their approvals in writing.
- Aggregate statistical data for three or more natural or legal persons may be disseminated but only provided these data no longer allow the identification of individual data. In particular, these data may not be disseminated if one natural or legal person represents over 80% of the total value or quantity of the units in the cell.

- Unique identifiers: personal code, fiscal code, name and address, locality and other information which cannot be depersonalized (made anonymous) will not be included in the individual data files for public use.

For scientific purposes access to the individual statistical data may be granted provided that all of the protection measures regarding the preservation of the confidentiality of statistical data have been taken and subject to:

- approval of the NIS management and endorsement by the Committee of Statistical Confidentiality, established through an Order issued by the NIS President;
- approval of the management of the Regional/County Statistical Office and endorsement by the Committee of Statistical Confidentiality, established at territorial level by decision of the executive manager of the Regional/County Statistical Office.

References

Eurostat (1999) *Manual on Disclosure Control Methods*

Brackstone, G. and White, P. (2002). *Data stewardship at Statistics Canada*. Proceedings of the Social Statistics Section, American Statistical Association, 284-293.

Eurostat (1996). *Manual on Disclosure Control Methods*. Luxembourg: Office for Official Publications of the European Communities.

Hundepool, A., van de Wetering, A., de Wolf, P.-P., Giessing, S., Fischetti, M., Salazar, J.-J. and Caprara, A. (2002). *ARGUS user manual 2.1*. Statistics Netherlands, Voorburg.

National Institute of Statistics *Rules on the Confidentiality of Statistical Data*

National Institute of Statistics *Policy for the Protection of Classified Information*

3.14 Data Analysis and Presentation

Scope and purpose

Data analysis is transforms data into information. Analytical results underscore the usefulness of data sources and shed light on issues. They are all the more significant when, for confidentiality reasons, it is not possible to release micro-data. They are typically presented in the form of commentaries attached to outputs and separately published articles.

The basic steps in analysis comprise identifying the issues, asking meaningful questions, developing answers to the questions through examination and interpretation of data, and communicating the messages to the users.

Analysis plays a role in survey development, review and enhancement by helping to identify problems and potential improvements. It facilitates effective extraction of information from previous surveys and pilot studies, planning new statistical activities, providing information on data gaps, for designing surveys, and formulating quality objectives.

Presentation of data and accompanying commentaries and of associated analytical articles requires consideration of how best to communicate with the users.

Principles

Analysis outputs may be categorized into two general types:

- *descriptive* results, relating to the survey population at the time that the data were collected, and
- *studies* that go beyond the actual population surveyed.

Commentaries attached to data outputs are restricted to descriptive results. The NIS mandate does not presently support analytical studies that interpret data output. Such studies may be conducted by independent researchers with NIS guidance.

Any conclusions presented in commentary or an analytical study, especially those that can impact on public policy, must be supported by the data being analysed.

Presentation standards should be developed and followed.

Guidelines

Preparation of Analytical Articles

- Identify the issues of concern (both current and those likely to emerge in the future), the audience, the sources of information and the analysis methods.
- Ensure that the data are appropriate for the analysis, for example that the frame population approximates the target population sufficiently closely, that the concepts and definitions are relevant, that the longitudinal or cross-sectional nature of the survey is appropriate, and that the data accuracy is sufficient for meaningful results.
- If more than one data source is being used for the analysis, investigate whether the sources are consistent and how they may be appropriately combined.
- Consider how unit and item non-response should be handled, and whether or not imputed values should be included.
- When making comparisons between two groups of individuals, businesses, or other units, control for extraneous factors. If significant differences between the groups are found, consider alternative plausible explanations.
- Since the analyses are based on observations rather than the results of a controlled experiment, avoid drawing conclusions concerning causality.
- Do not focus on short term trends without inspecting them in light of medium and long term trends. Frequently, short term trends are merely minor fluctuations around more important, longer term trends.
- Where possible, avoid arbitrary time reference points, such as the change from last year to this year. Instead, use meaningful points of reference, such as the last major turning point for economic data, generation-to-generation differences for demographic statistics, and legislative changes for social statistics.

Presentation of Analytical Articles

- Prior to writing an analytical article, prepare an outline of the article indicating the issues being addressed, the data being used, the analytical methods, the results to be highlighted, and the most interesting findings.
- Focus the article on important indicators and topics. Trying to be too comprehensive tends to interfere with a strong story line.
- Arrange ideas in a logical order and in order of relevance or importance. Use headings, sub-headings and sidebars to strengthen the organisation of the article.
- Keep the language as simple as the subject permits. Depending on the targeted audience, some loss of precision may be an acceptable trade-off for more readable text.
- Use graphs in addition to text and tables to communicate the message. Use headings that capture the meaning for example “Women’s earnings still trail men’s”, in preference to traditional titles like “Income by age and sex”.
- Aim to help readers understand the information in the tables and charts by discussing it in the text.

- When tables are used, take care that the overall format contributes to data clarity and minimises the possibility of misinterpretation. This includes appropriate spacing, wording, placement and appearance of titles, row and column headings and footnotes.
- Explain rounding practices or procedures. In the presentation of rounded data, do not use more significant digits than are consistent with the accuracy of the data.
- When presenting details about rates, be careful to distinguish between percentage change and change in percentage points. Define the base used for rates.
- Ensure that all references are accurate and are referred to in the text.
- Check for errors - ensure the consistency of figures used in the text, tables and charts and the accuracy of external data.
- Ensure that the intentions stated in the introduction are fulfilled and that the conclusions are consistent with the evidence.
- Have the article reviewed by at least two other persons.
- Prepare a verbal presentation of the analysis, and give it to colleagues before giving it to an external audience.

Commentaries

The guidelines above are equally applicable to commentaries attached to data outputs with the condition that the analysis is less in depth and is focused on summarising the main results rather than on exploring issues.

References

Chambers, R.L. and Skinner, C.J. (eds.) (2003). *Analysis of Survey Data*. Wiley, Chichester.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Duxbury Press.

3.15 Data Dissemination

Scope and purpose

Dissemination means data release to users. On line release mechanisms include the Internet, statistical databases and web files, television and radio interviews, and press conferences. Offline mechanisms include paper publications, CD-ROMs, press releases, fax, electronic mail, answers to ad-hoc requests, and presentations at events.

Principles

Disseminated outputs must be accurate, complete, accessible and appropriately priced, understandable, usable, timely and meet confidentiality requirements. Dissemination managers must take account of user needs and expectations based on user feedback, product testing and other marketing activities. They must exploit new technology to best enable users to access and process data easily and efficiently.

The dates and times when data will be released should be pre-announced. Users should be given equal treatment in terms of data access. Data should be published on the web, or sent electronically, or released in paper copies, at exactly the same time and in accordance with the pre-announcements.

Outputs must be accompanied by *metadata* that enable users to determine how appropriate the data are for the particular uses they have in mind. They also increase the confidence that users have in the data. The metadata provided should include concepts and definitions, targeted and sampled populations, sample design, questionnaire, data collection imputation and estimation methods, response rates, possible errors and exceptional circumstances. The

detail included should be matched to user needs. This may entail making summary metadata visible initially with links to more detailed metadata for users who need them.

Guidelines

- Ensure the data are checked and validated by a unit other than the producing unit. The checking unit should have proven experience.
- Develop and use a common layout and format for all products within the same category or theme and have a standard for the agency as a whole.
- Ensure that data released are exactly the same as are held in the internal databases. It should always be possible to reproduce the outputs exactly.
- Test every electronic product before release to ensure that it is performing as users would expect.
- Provide metadata containing appropriate descriptions of quality and methodology.
- Provide data quality indicators and tools to help users evaluate the data, for example CV look-up tables, sampling variance programs.
- Clearly state the rules, conditions and copyright terms for using or re-disseminating the statistical information.
- Accompany each output with a contact name, phone number and e-mail address.
- Make maximum use of generalised data and metadata dissemination systems.
- Ensure staff involved in data dissemination processes are well trained.
- Analyse and make use of dissemination standards and best practices from other EU countries.

References

National Institute of Statistics Romania (2006). *Policy on Data Dissemination*.

3.16 Documentation

Scope and purpose

Documentation here refers to a description of the survey/ statistical process, including the concepts and definitions, the methods used to produce the data, the outputs and the factors affecting quality. During survey development and implementation, documentation is a means of communication, including not only what decisions were made, but also why they were made. Documentation supports subsequent enhancement of the survey in the case of a repeating survey, and development of similar surveys. It also provides the basis for telling users about survey quality and methodology. Documentation may be aimed at any of several different groups, including managers, technical staff, planners of other surveys, and users.

Principles

The goal is to provide a complete, unambiguous and multi-purpose record of the survey and its outputs that is up-to-date, well organized, easily retrievable, concise and precise. The appropriate level of detail depends on the intended audience, data sources, sample design, collection and dissemination methods, users and uses of the data, and total budget.

Effective access and presentation is an important. Documentation may be multi-media, for example paper and electronic. Different documents may be prepared for different audiences and purposes.

There are numerous content standards of which the most important in the NIS context are the NIS own *Rules Regarding the Drafting and Editing of Survey Methodology* and *Rules Regarding the Drafting and Editing of Statistical Work* prepared by the Quality Unit, and the *Euro SDMX Metadata Structure (ESMS)*.

The ESMS contains a description and representation of the metadata to be used in documenting statistical data and in assessing data production processes and outputs. The ESMS concepts used are compatible with those of the *Statistical Data and Metadata Exchange (SDMX)* standard and with the common terminology in the *SDMX Metadata Common Vocabulary*. The ESMS will be implemented at Eurostat and national levels.

Guidelines

- Prepare survey documentation at level of detail appropriate for intended audiences.
- Indicate any changes in objectives, timeliness, frequency, and data quality targets that occurred during survey development, for example due to budgetary constraints, perceived feasibility, results of new pilot studies, or new technology.
- Indicate use of international and national standards and highlight deviations.
- Indicate roles of advisory committees and users in survey development and operations.
- Prepare versions of methodology documentation emphasising different aspects for different readers. Provide a methodological overview for senior managers and users. Provide a comprehensive and consolidated account of procedures and technical issues for professionals.
- Provide documents to accompany products and subject them to extensive review by peers and managers to ensure relevance, accuracy and readability
- Follow documentation revision policies.

References

Rules Regarding the Drafting and Editing of Survey Methodology

Rules Regarding the Drafting and Editing of Statistical Work

Euro SDMX Metadata Structure (ESMS).

Statistical Data and Metadata Exchange (SDMX).

Metadata Common Vocabulary.

3.17 Administrative Data Use

Scope and purpose

Administrative data are data collected during the course of administering government programs. Administrative procedures include collecting income and VAT taxes, regulating the flow of goods and people across borders, recording births, marriages and deaths, administering pensions, providing medical services. Administrative data are collected in accordance with relevant legislation. They are of two basic types:

- registration data, containing the details of all the units subject to the administrative procedure; and
- transaction data, containing details of the interactions of those units with the administrative procedure.

Each data record refers to an individual – person, tax payer, pensioner, etc - and the identity of the unit corresponding to a given record is crucial. By contrast, in the case of statistical data, the identity of individuals is of no intrinsic interest once the data have been collected.

Some administrative data are directly tabulated and disseminated by the NIS. For example shipments data are obtained from Customs, processed and disseminated as external trade data.

In addition to direct tabulation and dissemination, administrative data may be used:

- to construct and maintain the business register;
- to replace of data collection partially or entirely, for example use of taxation data for small businesses in place of direct collection of income and expenditure data;
- to assist in editing and imputation;
- as auxiliary information in estimation - for calibration benchmarking and/or calendarisation; and
- for survey evaluation, in particular *data confrontation* and comparison of survey and administrative data.

Principles

Advantages

The advantages of using administrative data are, first, by law, they are provided free of charge by the administrative agency responsible for their collection. Second, they impose no additional statistical respondent burden on the individuals to which the data refer. Third, modern technology enables the NIS to readily access and process large administrative datasets.

The disadvantages of administrative data stem from the fact that they are collected for administrative rather than statistical purposes. Usually, the potential statistical uses of the administrative data are unknown or considered unimportant at the time the administrative program is put in place. This means that the NIS has little or no direct control over coverage, content, accuracy or timing. Further more changes in legislation can causes changes in the data, again outside NIS control.

In summary, at a time when data demands are growing, there are severe budgetary constraints and growing concerns about response rates and respondent burden, the use of administrative data is an attractive complement and/or alternative to direct data collection. Thus, it is NIS policy to use administrative data whenever they present a cost beneficial alternative, or addition, to direct data collection.

As with any data acquisition program, the decisions to use an administrative data source must be preceded by an assessment of the data in terms of their coverage, content, concepts and definitions, frequency, timeliness, and quality assurance procedures, as well as the likely stability of the administrative program over time, and the costs of acquiring and processing the data.

In processing administrative data the NIS has to undertake virtually all the same sub-processes as for statistical data discussed in previous sections. Thus, the guidelines in earlier sections are for the most part equally applicable to administrative records. Even sampling and data capture guidelines may be relevant if administrative data exist only on paper and have to be sampled, coded and captured. However, there are some additional considerations, which are the subject of this section.

First, the use of administrative data may raise privacy concerns. These are even more important when the administrative data are linked to other sources of data. Thus such linkage is usually carefully controlled.

Second, it is not easy to combine data from an administrative source with data from another source in the absence of a common identification system covering both sources. *Record linkage* techniques are required. If the purpose of linking the sources is frame creation and

maintenance, edit and imputation or weighting, exact matching is essential. If, on the other hand, sources are being linked for data analysis purposes *statistical matching*, i.e., matching of records with similar statistical properties may be sufficient.

Guidelines

- Build and maintain close ties with the custodians of the administrative sources. Liaison is necessary not only in negotiating use of administrative data, but also in ensuring that the NIS is not caught by surprise by changes to administrative procedures, and can even influence them. Feedback of problems and suggestions to the data custodians can lead to strengthening of the administrative sources. Occasionally, the statistical office may be able to help build statistical requirements into new programs before they commence. Such opportunities are rare, but when they happen, the eventual value to the NIS far outweighs the time and resources required for collaboration.
- Understand the context under which the administrative organization created the administrative program (e.g., legislation, objectives, and needs). It has a profound impact on the coverage, contents, concepts and definitions, frequency and timeliness, quality and stability over time.
- Study each data item in the administrative source that is to be used for statistical purposes. Investigate its quality. Understand the concepts, definitions and procedures underlying its collection and processing. Some items might be of very poor quality and not fit for use. For example, the quality of classification coding (e.g., occupation, industrial activity, geography) may be insufficient for statistical purposes or may severely curtail use. In such cases consider the possibility of providing NIS staff or resources to improve quality.
- Choose methodology for the survey that is appropriate for the particular administrative source.
- Like data collected by survey, administrative data are also subject to partial and total non-response. In some instances, the lack of timeliness in obtaining the data effectively results non-response. Thus, imputation and weight adjustment procedures must be developed.
- Imputation and/or transformation may also be required in cases where (some) units report the data at a different frequency than is required for survey purposes.
- The activity statuses of units according to an administrative source are frequently different from those required for statistical purposes. For example the tax agency will retain a bankrupt business in its records so long as it owes tax whereas for the statistical office the business is inactive. Thus, special attention must be given to the identification of administrative units that are apparently active but inactive from a statistical perspective. For example not submitting value added tax for 18 months may be considered evidence of being inactive.
- Also, keep in mind that there may be significant measurement errors if the administrative data have financial implications. For example, income may be underestimated and/or expenditures overestimated to reduce tax liability.
- Consider privacy implications of the publication of information from administrative records. Although the NIS has the authority to access administrative records for statistical purposes, this use may not have been foreseen by the original suppliers of information. Therefore, survey managers should be prepared to explain and justify the public value and innocuous nature of use for statistical purposes.
- Implement continuous or periodic assessment of incoming data quality. Assurance that data quality is being maintained is important because the NIS does not control the data collection process. Assessment may include use of statistical quality control methods, comparisons with other sources and sample follow-up studies.

- The privacy concerns that may arise with use of a single administrative record source are multiplied when linkage is made to other sources. A record linkage policy must be designed to ensure that the value of linkage outweighs any intrusion on privacy that it may entail.
- When data are being brought together for two or more sources without a common identification scheme, consider statistical matching methods when exact matching is not required.

References

Cox, L.H. and Boruch, R.F. (1988). *Record Linkage, Privacy and Statistical Policy*. Journal of Official Statistics, 4, 3-16.

Eurostat (2001) *Use of Administrative Sources for Business Statistics Purposes*

Hidiroglou, M.A., Latouche, M.J., Armstrong, B., and Gossen, M. (1995). *Improving Survey*

Information Using Administrative Records: The Case of the Canadian Employment Survey. Proceedings of the Annual Research Conference, U.S. Bureau of the Census, 171-197.

Michaud, S., Dolson, D., Adams, D., and Renaud, M. (1995). *Combining Administrative and Survey Data to Reduce Respondent Burden in Longitudinal Surveys*. Proceedings of the Section on Survey Research Methods, American Statistical Association, 11-20.

4 Quality Management Framework

Introductory Remarks

This chapter outlines the *Quality Management Framework* (QMF) within which the quality guidelines described in Chapter 3 are applied. Whereas Chapter 3 refers to surveys and is aimed at survey managers, Chapter 4 refers to the agency as a whole and is aimed at senior managers and staff with special responsibility for promoting quality management within the organisation.

Quality management involves identification and pursuit of quality objectives in the light of user demands for more information and other quality considerations, taking into account financial and human resources, and the burden on respondents. Thus, the QMF comprises many mechanisms at various levels throughout the NIS. Its effectiveness depends on the professionalism and motivation of the staff and the synergy resulting from the many players operating collectively with a common set of quality goals in mind.

The QMF is presented in the context of the *output quality components* defined in Chapter 2, with sections on *Managing Relevance, Accuracy, Timeliness, Accessibility, Comparability and Coherence*, together with two cross-cutting sections on *Quality Reporting* and *Managing the Institutional Environment*.

4.1 Managing Relevance

Relevance is the degree to which statistical outputs meet current and potential user needs. It depends on whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.,) reflect user needs.

The NIS programme and outputs must properly and continuously reflect the most important data needs as these evolve over time. Thus management of relevance has two aspects – determining statistical needs and adjusting the NIS programme to meet these needs.

Determining User Needs

ESS requirements are specified in EU regulations. They are mandatory. Furthermore, quality reports required by the regulations typically require evidence that the required data are being produced. Additional needs may come from other international agencies in particular the IMF, OECD, UNDP, FAO, etc.

National demands are identified through liaison with major users, for the most part government departments. There is also some feedback from other users on existing outputs collected by the Dissemination Bureau. Researchers and analysts may indicate data gaps and limitations through published articles. Comments and criticisms from media may also be informative.

Account also has to be taken of the needs of internal programmes, such as the System of National Accounts, that draw data from a range of surveys and integrate and reconcile them.

The processes that determine data needs also result in the monitoring of other quality components and identification of the corresponding user requirements.

Adjusting the Programme

Adjustments to the programme are made annually, through the definition and implementation of the National Annual Statistical Work Programme. The Work Programme defines outputs, tasks and timelines for NIS statistical activities funded by the government. Each year the Work Programme for the following year is put together based on that for the current year but adjusted to take into account changing needs. The Work Programme is then circulated to the primary users – government departments and agencies – for comments. Based on feedback

received, and in the light of current policy priorities, NIS senior management finalises the Work Programme.

In practice, there are heavy constraints on program adjustments. First, a large proportion of the NIS's budgetary resources are devoted to ongoing surveys and statistical processes that are *non-discretionary* in the sense that they are dictated by EU regulations or provide such basic information on Romanian society and economy that they cannot be significantly changed. Second, it is extremely difficult to discontinue, or even cut back, an existing survey because the users will make vociferous objections. Third, there are essentially no prospects of securing additional government funding for new programmes unless such funding is actually built into accompanying legislation.

Thus, contractions to the programme are minimal and extensions to meet new data needs are generally very limited and undertaken only by diverting resources from ongoing activities.

Each year some statistical activities are funded from other sources, for example by the EU, the World Bank, and/or other international and national partners. However, these are typically aimed at improvements to the statistical processes.

Improvement Considerations

The current procedures for managing relevance could be improved in three ways. First, further extent of usage of multi-annual planning process would give more scope for entertaining programme changes.

Second, regular self administered reviews of all surveys would provide more feedback on public sensitivities and relevance at survey level, and the opportunities for changes or cuts to question modules or sample sizes. Such reviews are envisaged beginning in 2010, as further described in the section on quality reporting.

Third, options for releasing resources to undertake new surveys or expansion of existing ones could be explored. These included obtaining a better understanding of costs of individual surveys so that the opportunities for efficiencies could be more easily identified, and measuring respondent burden as a lever for reducing questionnaire size and sample size.

4.2 Managing Accuracy

*The **accuracy** of statistical outputs in the general statistical sense is the degree of closeness of estimates to the true values.*

Design Aspects

The accuracy, timeliness and coherence achieved in survey outputs depend heavily upon the design methodology, including the quality assurance processes put in place to identify and control potential errors. There are two aspects: the methodological capacity and prevailing policies and procedures; and the application of this capacity, policies and procedures to individual surveys.

There is currently no methodology unit or primary centre of expertise. Responsibility for sampling and estimation is invested in three units handling business, household and agricultural surveys respectively. All other methodological matters, for example questionnaire design, editing procedures, imputation procedures, outlier detection, are handled by the statisticians responsible for the individual surveys. The Quality Unit takes responsibility for promoting quality practices, for example quality reporting and the production and use of these quality guidelines.

As regards methodological procedures, these guidelines describe the specific practices, methods and considerations that should be taken into account in designing programs, and indicate where formal standards exist. Apart from this document there are few general NIS

methodological procedures, for example questionnaire design and layout guidelines. It is standard practice to pre-test and pilot test all new questionnaires prior to use in production. Staff rely on ESS and international literature and experts for guidance, for example the ESS Guidelines on Seasonal Adjustment.

As regards application of methodology, decisions are largely left to the individual survey managers and their staff. In addition to deciding what constitute acceptable accuracy targets on the basis of their knowledge of user requirements, budget and operations, they determine the specific methods to be applied.

Implementation Aspects

Survey accuracy depends planning, operational, procedures, supervision, respondent relations, interviews, and training. Quality and performance monitoring mechanisms should be built into survey processes as part of design. Two types of information are required. First information is needed to monitor and correct immediately any operational problems. Second information is needed to assess whether the survey was implemented as planned, whether some aspects of the design were problematic, and what lessons were learned from the operational standpoint to aid future designs.

Assessment Aspects

Accuracy assessment must be considered at the design stage as measurements need to be obtained as the survey is taking place. The extent and sophistication of measurements depend on survey size and budget, and their significance to users. Informing users of data quality requires assessment of, at least, the following aspects: survey coverage, sampling error, non-response rates, and descriptions of any serious problems or issues.

Improvement Considerations

Possibilities for further improvements in managing accuracy including reinforcing expertise in sampling and estimation, questionnaire design, editing, imputation, and analysis, building quality and performance monitoring mechanisms into survey processes, introducing the programme of user oriented quality reports (as further discussed in later section, encouraging use of the survey (DESAP) checklist, and use of new technologies and innovations (with adequate testing).

4.3 Managing Timeliness and Punctuality

*The **timeliness** of statistical outputs is the length of time between the event or phenomenon they describe and their availability. Information that is available to users well within the period during which it remains useful for its main purposes is considered to be timely.*

***Punctuality** is the time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by regulations or previously agreed among partners.*

Timeliness and punctuality are constantly monitored over time and across programs through the weekly exception reports to senior management.

The choice of timeframe for a survey, including release dates is a design decision, usually depending on trade-offs with accuracy and with cost. There are two separate catalogues of outputs produced towards the end of the year and visible via the NIS web site. They specify the release dates of electronic and printed publications and of press releases for the following year.

For some programs, the release of preliminary data followed by revised and final figures is used as a strategy for making data timelier. In such cases, the tracking of the size and direction of revisions can serve to assess the appropriateness of the chosen timeliness-accuracy trade-off. It

also provides a basis for recognizing any persistent or predictable biases in preliminary data that could be removed through estimation.

Improvements in timeliness can be expected as new technologies are developed and as uses of data change.

4.4 Managing Accessibility

The **accessibility** of statistical outputs is the measure of the ease with which users can obtain the data. It is determined by the physical conditions by means of which users obtain data: where to go, how to order, delivery time, pricing policy, marketing conditions (copyright, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, Internet, etc.)

At the beginning of a new survey or following major survey redesign, the survey manager prepares an initial dissemination proposal. This includes what data to disseminate, how and when based on known user demands and further discussions with users. The proposal is reviewed by senior management, modified in accordance with their decisions and then implemented by the Dissemination Bureau.

The primary NIS dissemination vehicles are: paper publications, electronic publications, TEMPO online database, the public library, dissemination bureau.

There are a variety of options that are open to survey managers to make their data more accessible for analytical purposes, including:

- production of public-use micro-data files;
- provision of a custom data provision service;
- contracts with external researchers.

4.5 Managing Clarity

The **clarity** of statistical outputs is the measure of the ease with which users can understand the data. It is determined by the information environment within which the data are presented, whether the data are accompanied with appropriate metadata, whether use is made of illustrations such as graphs and maps, whether information on data accuracy are available (including any limitations on use) and the extent to which additional assistance is provided by the producer.

Metadata

Providing sufficient information to allow users to properly interpret output data is an NIS responsibility and is primarily about the provision of appropriate metadata. The information needed to understand statistical data falls under three broad headings:

- concepts, variables and classifications that underlie the data (conceptual metadata);
- methodology used to collect and compile the data (procedural metadata); and
- indicators of the accuracy of the data (operational metadata).

The commentaries that accompany output data are a further aid to users.

The primary vehicle for disseminating metadata is the Metadata Database, which can be viewed via the Internet. It is not currently linked at data item level to the TEMPO database and this is an area for improvement. Also there are several separate in-house sources of metadata that should be consolidated within the Metadata Database to enrich its content and inhibit variations.

Managing Misinterpretations

Media are monitored every day and, in event of a problem, the Press Relations Unit takes action. Serious misinterpretations of data by the media or in the public arena are immediately challenged publicly by the NIS.

4.6 Managing Coherence and Comparability

The **coherence** of two or more statistical outputs refers to the degree to which the statistical processes by which they were generated used the same concepts - classifications, definitions, and target populations – and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly. **Comparability** is a special case of coherence and refers to the case where the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains.

Three complementary approaches are used in managing coherence.

The first approach is the development and use of standard concepts, definitions and classification systems for all major indicators as well as implementation of international standard frameworks, for example, the System of National Accounts, where these exist.

The second approach is to ensure that the measurement methods themselves do not introduce inconsistencies, in particular by use of:

- the business register as the frame for all business surveys;
- the farm register as the frame for all agricultural surveys;
- the master sample of households for all household surveys
- harmonised methodologies and systems; and
- quality guidelines and international standards.

Areas for improvement include the documentation and use of commonly formulated questions, the documentation and use of questionnaire design and layout standards.

The third approach involves analysis of the data themselves and comparison and integration of data from different sources or over time. The aim is to recognize situations where variation or inconsistency exceeds levels implied by the expected accuracy of the data. Integration of data in the national accounts is an example. Also, measures of coherence are required in certain quality reports prepared for Eurostat in accordance with EU regulations. For example a report on the Labour Cost Survey summarised the coherence of the survey outputs with the National Accounts, the Labour Force, and the Structural Business Survey.

Users are also an important source of information on coherence, or more particularly lack of it.

4.7 Managing Quality Reporting

Scope of Quality Reporting

The current and proposed spectrum of quality assessment methods may be described in terms of three levels, as indicated in Figure 1.

Level 1: Output quality indicators, process quality and performance indicators, quality reports, and user feedback/survey results.

Data obtained from measurements of processes and outputs are selected and structured into quality and performance indicators, which are themselves grouped in quality reports. User surveys are conducted to measure user perceptions of product quality. Quality reports are prepared to meet legislated or other specific information requirements.

Level 2: Self-assessments; comprehensive reports and peer reviews

Based on Level 1 information and further analysis, surveys and their products will be assessed against a comprehensive internal standard developed by the Quality Unit. Such assessments will give a more comprehensive picture than Level 1 assessments and, being conducted according to a common format, they will enable some comparisons across surveys.

Level 3: Labelling and certification.

This level has yet to be implemented. Labelling is aimed at providing highly condensed quality information regarding the statistics released for the benefit of users. It will involve assigning a label to the outputs of those surveys that meet specific NIS quality requirements. Certification is similar to labelling, but is against an external standard such as EFQM Excellence Model and refers to the NIS as a whole, rather than to specific products. There are presently no plans for labelling or certification.

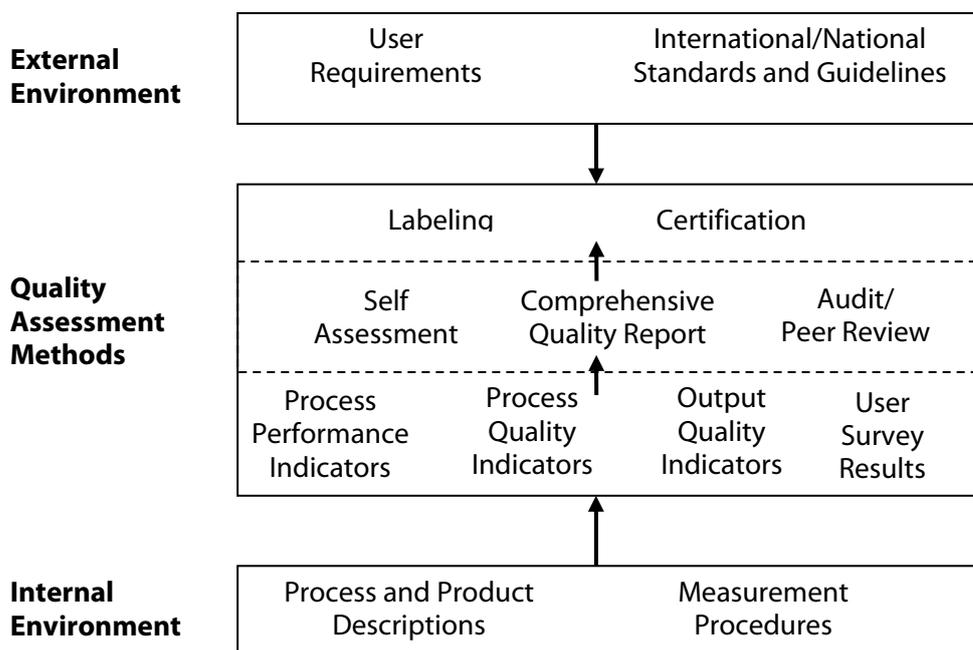


Figure 1: Quality Assessment Methods

User Oriented Quality Reports

As previously indicated, a programme of user oriented quality reports will begin in 2010. In principle, every survey will be required to prepare a report every year. However, in the first year, surveys compiling quality reports for Eurostat will be excluded. The reports will include the ESS standard set of quality and performance indicators

A report template and instructions for its completion have been developed.

Checklists

The NIS is promoting use the *European Self Assessment Checklist for Survey Managers* (DESAP) as a way of conducting of quick but systematic quality assessment of a survey and a means of identifying potential improvements.

4.8 Managing the Institutional Environment

The management of quality takes place within the NIS institutional environment. Measures in place to create and support an environment and culture that recognizes the importance of quality include the following:

- A program of recruitment and development for major occupational groups.
- An overall training and development framework.
- The National Training Centre.
- A staff opinion survey was conducted in 2004 and another is scheduled for 2011.
- Establishing specific areas of expertise and oversight committees, for example in sampling and estimation, and in IT applications.
- Protocols and conventions with other government departments and agencies that are either providers of administrative data, users of statistical data or both.
- Procedures to handle respondent complaints.
- Establishment of respondent burden measurements for business surveys.

There is considerable scope for improvements for example (1) establishing and reinforcing areas of methodological expertise, (2) encouraging data analysis both to better inform users via commentaries on data outputs and to detect data problems and deficiencies, (3) extending respondent burden measurements to all statistical surveys.

References

Eurostat (2002) *Business Methods Your Guide to Business Statistics Methodology*.

Eurostat (2000) *Handbook on Design and Implementation of Business Surveys*.

Fellegi, I. (1996). *Characteristics of an effective statistical system*. International Statistical Review, 64, 165-197

Trewin, D. (2002). *The importance of a quality culture*. Survey Methodology, 28, 125-133.

5 Appendix: European Statistical System Quality Policies

5.1 ESS Quality Declaration

The mission of the European Statistical System

“We provide the European Union and the World with high quality information on the economy and society at the European, national and regional levels and make the information available to everyone for decision-making purposes, research and debate.”

The vision of the European Statistical System

“The ESS will be a world leader in statistical information services and the most important information provider for the European Union and its member states. Based on scientific principles and methods, the ESS will offer and continuously improve a programme of harmonized European statistics that constitutes an essential basis for democratic processes and progress in society.”

User focus

We provide our users with products and services that meet their needs. The articulated and non-articulated needs, demands and expectations of external and internal users will guide the ESS, its members, their employees and operations.

Continuous improvement

The needs and demands of users will change as will the environment we operate in. Globalisation and advances in methods and technology will avail new possibilities. It is imperative that we actively strive to improve our work methods to take advantage of the new possibilities and to better meet the demands of our users.

Product quality commitment

We produce high quality statistical information according to scientific methods in accordance with objectivity and confidentiality. We provide information on the main quality characteristics of each product so that users are able to assess product quality.

Accessibility of information

We provide statistical results in a user-friendly and accessible form. Utilizing the possibilities of new media ensures easy access to the information. As far as possible, we will enhance user awareness of the strengths and limitations of the produced statistics. Consulting on how to use data is an integral part of dissemination.

Partnership with and beyond the European Statistical System

The cooperation between current and future members of the ESS as well as with other organisations will be encouraged. Only by working together, can we learn from others and gradually develop our system. The broad knowledge of staff and our users, suppliers, partners and other parties must be combined for us to excel in our purpose.

Respects for the needs of data providers

The suppliers of data for statistics – the respondents – are an especially important group with which a mutually rewarding partnership must be established. The producers of statistics should strive to always minimise the respondent burden, both the objective and the perceived burden.

Commitment of leadership

The leaders of the organisations in the ESS exercise a personal, active, and visible leadership to create and sustain a culture of quality. By providing a clear overall direction, prioritizing improvement activities and stimulating empowerment and innovation, leaders enable the staff to perform a successful job and to continuously strive for improvement.

Systematic quality management

We systematically and regularly identify strengths and weaknesses in all relevant areas to continuously identify and implement improvements where needed. A long-term strategic orientation is vital for the development of the ESS. The long-term effects in all situations must be considered with the more obvious short-term effects.

Effective and efficient processes

ESS activities should be seen as processes that create value for the users. We work efficiently to produce output with as little resources as possible and to prevent errors in the processes and products. The processes and their quality are continuously reviewed and improved.

Staff satisfaction and development

To attract and keep competent staff, it is vital to satisfy staff needs. The ESS members should treat their employees as the key resources they are.

5.2 European Statistics Code of Practice

For the national and community statistical authorities

Adopted by the Statistical Programme Committee on 24 February 2005 and promulgated in the Commission Recommendation of 25 May 2005 on the Independence, Integrity and Accountability of the National and Community Statistical Authorities

The European Statistics Code of Practice is based on 15 principles. Governance authorities and statistical authorities in the European Union commit themselves to adhering to the principles fixed in this code covering the institutional environment, statistical processes and outputs. A set of indicators of good practice for each of the 15 principles provides a reference for reviewing the implementation of the Code.

Institutional environment

Institutional and organisational factors have a significant influence on the effectiveness and credibility of a statistical authority producing and disseminating European statistics. The relevant issues are professional independence, mandate for data collection, adequacy of resources, quality commitment, statistical confidentiality, impartiality and objectivity.

PRINCIPLE 1: PROFESSIONAL INDEPENDENCE

The professional independence of statistical authorities from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, ensures the credibility of European statistics.

Indicators

- The independence of the statistical authority from political and other external interference in producing and disseminating official statistics is specified in law.

- The head of the statistical authority has sufficiently high hierarchical standing to ensure senior-level access to policy authorities and administrative public bodies. He/she should be of the highest professional calibre.
- The head of the statistical authority and, where appropriate, the heads of its statistical bodies have responsibility for ensuring that European statistics are produced and disseminated in an independent manner.
- The head of the statistical authority and, where appropriate, the heads of its statistical bodies have the sole responsibility for deciding on statistical methods, standards and procedures, and on the content and timing of statistical releases.
- The statistical work programmes are published, and periodic reports describe progress made.
- Statistical releases are clearly distinguished and issued separately from political/policy statements.
- The statistical authority, when appropriate, comments publicly on statistical issues, including criticisms and misuses of official statistics.

PRINCIPLE 2: MANDATE FOR DATA COLLECTION

Statistical authorities must have a clear legal mandate to collect information for European statistical purposes. Administrations, enterprises and households, and the public at large may be compelled by law to allow access to or deliver data for European statistical purposes at the request of statistical authorities.

Indicators

- The mandate to collect information for the production and dissemination of official statistics is specified in law.
- The statistical authority is allowed by national legislation to use administrative records for statistical purposes.
- On the basis of a legal act, the statistical authority may compel response to statistical surveys.

PRINCIPLE 3: ADEQUACY OF RESOURCES

The resources available to statistical authorities must be sufficient to meet European statistics requirements.

Indicators

- Staff, financial, and computing resources, adequate both in magnitude and in quality, are available to meet current European statistics needs.
- The scope, detail and cost of European statistics are commensurate with needs.
- Procedures exist to assess and justify demands for new European statistics against their cost.
- Procedures exist to assess the continuing need for all European statistics, to see if any can be discontinued or curtailed to free up resources.

PRINCIPLE 4: QUALITY COMMITMENT

All ESS members commit themselves to work and cooperate according to the principles fixed in the 'Quality declaration of the European statistical system'.

Indicators

- Product quality is regularly monitored according to the ESS quality components.
- Processes are in place to monitor the quality of the collection, processing and dissemination of statistics.
- Processes are in place to deal with quality considerations, including trade-offs within quality, and to guide planning for existing and emerging surveys.
- Quality guidelines are documented and staff are well trained. These guidelines are spelled out in writing and made known to the public.
- There is a regular and thorough review of the key statistical outputs using external experts where appropriate.

PRINCIPLE 5: STATISTICAL CONFIDENTIALITY

The privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes must be absolutely guaranteed.

Indicators

- Statistical confidentiality is guaranteed in law.
- Statistical authority staff sign legal confidentiality commitments on appointment.
- Substantial penalties are prescribed for any wilful breaches of statistical confidentiality.
- Instructions and guidelines are provided on the protection of statistical confidentiality in the production and dissemination processes. These guidelines are spelled out in writing and made known to the public.
- Physical and technological provisions are in place to protect the security and integrity of statistical databases.
- Strict protocols apply to external users accessing statistical microdata for research purposes.

PRINCIPLE 6: IMPARTIALITY AND OBJECTIVITY

Statistical authorities must produce and disseminate European statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably.

Indicators

- Statistics are compiled on an objective basis determined by statistical considerations.
- Choices of sources and statistical techniques are informed by statistical considerations.
- Errors discovered in published statistics are corrected at the earliest possible date and publicised.
- Information on the methods and procedures used by the statistical authority are publicly available.
- Statistical release dates and times are pre-announced.
- All users have equal access to statistical releases at the same time and any privileged pre-release access to any outside user is limited, controlled and publicised. In the event that leaks occur, pre-release arrangements should be revised so as to ensure impartiality.
- Statistical releases and statements made in press conferences are objective and non-partisan.

Statistical processes

European and other international standards, guidelines and good practices must be fully observed in the processes used by the statistical authorities to organise, collect, process and disseminate official statistics. The credibility of the statistics is enhanced by a reputation for good management and efficiency. The relevant aspects are sound methodology, appropriate statistical procedures, non-excessive burden on respondents and cost effectiveness.

PRINCIPLE 7: SOUND METHODOLOGY

Sound methodology must underpin quality statistics. This requires adequate tools, procedures and expertise.

Indicators

- The overall methodological framework of the statistical authority follows European and other international standards, guidelines and good practices.
- Procedures are in place to ensure that standard concepts, definitions and classifications are consistently applied throughout the statistical authority.
- The business register and the frame for population surveys are regularly evaluated and adjusted if necessary in order to ensure high quality.
- Detailed concordance exists between national classifications and sectorisation systems and the corresponding European systems.
- Graduates in the relevant academic disciplines are recruited.
- Staff attend international relevant training courses and conferences, and liaise with statistician colleagues at international level in order to learn from the best and to improve their expertise.
- Cooperation with the scientific community to improve methodology is organised and external reviews assess the quality and effectiveness of the methods implemented and promote better tools, when feasible.

PRINCIPLE 8: APPROPRIATE STATISTICAL PROCEDURES

Appropriate statistical procedures, implemented from data collection to data validation, must underpin quality statistics.

Indicators

- Where European statistics are based on administrative data, the definitions and concepts used for the administrative purpose must be a good approximation to those required for statistical purposes.
- In the case of statistical surveys, questionnaires are systematically tested prior to the data collection.
- Survey designs, sample selections, and sample weights are well based and regularly reviewed, revised or updated as required.
- Field operations, data entry, and coding are routinely monitored and revised as required.
- Appropriate editing and imputation computer systems are used and regularly reviewed, revised or updated as required.
- Revisions follow standard, well-established and transparent procedures.

PRINCIPLE 9: NON-EXCESSIVE BURDEN ON RESPONDENTS

The reporting burden should be proportionate to the needs of the users and should not be excessive for respondents. The statistical authority monitors the response burden and sets targets for its reduction over time.

Indicators

- The range and detail of European statistics demands is limited to what is absolutely necessary.
- The reporting burden is spread as widely as possible over survey populations through appropriate sampling techniques.
- The information sought from businesses is, as far as possible, readily available from their accounts and electronic means are used where possible to facilitate its return.
- Best estimates and approximations are accepted when exact details are not readily available.
- Administrative sources are used whenever possible to avoid duplicating requests for information.
- Data sharing within statistical authorities is generalised in order to avoid multiplication of surveys.

PRINCIPLE 10: COST EFFECTIVENESS

Resources must be effectively used.

Indicators

- Internal and independent external measures monitor the statistical authority's use of resources.
- Routine clerical operations (e.g. data capture, coding and validation) are automated to the extent possible.
- The productivity potential of information and communications technology is being optimised for data collection, processing and dissemination.
- Proactive efforts are being made to improve the statistical potential of administrative records and avoid costly direct surveys.

Statistical Output

Available statistics must meet users' needs. Statistics comply with European quality standards and serve the needs of European institutions, governments, research institutions, business concerns and the public generally. The important issues concern the extent to which the statistics are relevant, accurate and reliable, timely, coherent, comparable across regions and countries, and readily accessible by users.

PRINCIPLE 11: RELEVANCE

European statistics must meet the needs of users.

Indicators

- Processes are in place to consult users, monitor the relevance and practical utility of existing statistics in meeting their needs, and advise on their emerging needs and priorities.
- Priority needs are being met and reflected in the work programme.
- User satisfaction surveys are undertaken periodically.

PRINCIPLE 12: ACCURACY AND RELIABILITY

European statistics must accurately and reliably portray reality.

Indicators

- Source data, intermediate results and statistical outputs are assessed and validated.
- Sampling errors and non-sampling errors are measured and systematically documented according to the framework of the ESS quality components.
- Studies and analyses of revisions are carried out routinely and used internally to inform statistical processes.

PRINCIPLE 13: TIMELINESS AND PUNCTUALITY

European statistics must be disseminated in a timely and punctual manner.

Indicators

- Timeliness meets the highest European and international dissemination standards.
- A standard daily time is set for the release of European statistics.
- Periodicity of European statistics takes account of user requirements as much as possible.
- Any divergence from the dissemination time schedule is publicised in advance, explained and a new release date set.
- Preliminary results of acceptable aggregate quality can be disseminated when considered useful.

PRINCIPLE 14: COHERENCE AND COMPARABILITY

European statistics should be consistent internally, over time and comparable between regions and countries; it should be possible to combine and make joint use of related data from different sources.

Indicators

- Statistics are internally coherent and consistent (e.g. arithmetic and accounting identities observed).
- Statistics are coherent or reconcilable over a reasonable period of time.
- Statistics are compiled on the basis of common standards with respect to scope, definitions, units and classifications in the different surveys and sources.
- Statistics from the different surveys and sources are compared and reconciled.
- Cross-national comparability of the data is ensured through periodical exchanges between the European statistical system and other statistical systems; methodological studies are carried out in close cooperation between the Member States and Eurostat.

PRINCIPLE 15: ACCESSIBILITY AND CLARITY

European statistics should be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

Indicators

- Statistics are presented in a form that facilitates proper interpretation and meaningful comparisons.

- Dissemination services use modern information and communication technology and, if appropriate, traditional hard copy.
- Custom-designed analyses are provided when feasible and are made public.
- Access to microdata can be allowed for research purposes. This access is subject to strict protocols.
- Metadata are documented according to standardised metadata systems.
- Users are kept informed on the methodology of statistical processes and the quality of statistical outputs with respect to the ESS quality criteria.

Reproducerea conținutului acestei publicații, integrală sau parțială, în formă originală sau modificată, precum și stocarea într-un sistem de regăsire sau transmiterea sub orice formă și prin orice mijloace sunt interzise fără autorizarea scrisă a Institutului Național de Statistică.

Utilizarea conținutului acestei publicații, cu titlu explicativ sau justificativ, în articole, studii, cărți, este autorizată numai cu indicarea clară și precisă a sursei.

Reproducing the content of this publication, in its entirety or partially, in its original form or in a modified form, and also its keeping in a retrieval system or its transmission under any form and by any means are forbidden unless there is a written authorization in this respect from the National Institute for Statistics.

The use of the content of this publication, as an explanation or justification, in articles, studies or books is authorized only if the source is clearly and precisely mentioned.