

 QUALITY
 MANAGEMENT
 FRAMEWORK
2025



QUALITY MANAGEMENT FRAMEWORK 2025

Published by
National Statistics Office
Lascaris, Valletta VLT 2000
Malta
Tel.: (+356) 2599 7000
e-mail: nso@gov.mt
website: <http://www.nso.gov.mt>

CIP Data

Quality Management Framework 2025 –
Valletta: National Statistics Office, Malta, 2025
122p.

Print: ISBN: 978-9918-605-48-4
PDF: ISBN: 978-9918-605-49-1

NSO publications are available from:

Communication and Dissemination Unit
National Statistics Office
Lascaris
Valletta VLT 2000
Malta
Tel: (+356) 2599 7219
Email: nso@gov.mt

For further information about the content of this publication, please contact:

Methodology and Quality Unit
National Statistics Office
Lascaris
Valletta
Tel: (+356) 2599 7309
Email: methodology.nso@gov.mt



FOREWORD



ETIENNE CARUANA

Director General

March 2025

This document outlines the first Quality Management Framework (QMF) of the NSO which establishes guidelines and procedures to ensure and enhance the quality of statistical data produced by the Office. It covers the entire statistical production process, from initial planning to final dissemination, emphasising the importance of accurate and reliable data for informed decision-making and policy formulation.

The QMF highlights the need for statistical surveys to be precise and well-executed, with continuous monitoring and feedback loops to identify and address potential issues. Integrating administrative data is promoted to improve quality and reduce the burden on respondents. Data confidentiality and security are also prioritised, ensuring privacy protection and statistical-only usage.

A commitment to standardised concepts, definitions, and classifications ensures consistency and comparability of data. Recognising that statistical processes are subject to errors, the document provides guidance on assessing and mitigating them to maintain the reliability of statistical estimates.

Metadata plays a key role in documenting statistical processes and enhancing user understanding. This is also complemented by clear and transparent data presentation which is key, adhering to established guidelines for text, tables, and charts.

The QMF reflects NSO’s commitment to quality, which is one of the main pillars that that distinguishes official statistics from other forms of statistics. By adhering to these principles, NSO can enhance the quality, relevance, and reliability of its statistical outputs, fostering public trust and supporting evidence-based decision-making. I trust these guidelines can also prove useful to our statistical partners that either produce their own statistics or contribute to the production of official statistics.

As a living document, the QMF will be continuously updated to reflect emerging best practices, evolving methodologies, and user needs. Regular revisions and public sharing will ensure transparency, adaptability, and alignment with the highest statistical standards. Through this dynamic approach, the NSO reaffirms its dedication to producing reliable, relevant, and high-quality data that serve the public interest.

I would like to thank the Methodology and Quality Unit within the Data Resources, IT and Methodology for leading this project. My appreciation also goes to the several other Units that were either consulted or contributed directly to this publication.





TABLE OF CONTENTS

	Page
01. Statistical Surveys	9
02. Accessing, Using and Managing the Use of Administrative Data	13
03. Collaboration, Data Sharing, and Relationships with Data Providers	17
04. Statistical Confidentiality	21
05. Sampling Frames and Coverage	29
06. Questionnaire Design and Testing	33
07. Interviewer Recruitment, Training and Assignment	39
08. Methods of Data Collection	43
09. Response Burden	47
10. Consistent Application of Standards on Concepts, Definitions and Classifications	51
11. Statistical Data Editing and Imputation	55
12. Statistical Estimation and Analysis	63
13. Indices and Indicators	73
14. Weight Construction and Adjustment for Non-Response	79
15. Consistency Checks	87
16. Time Series and Seasonal Adjustment	91
17. Assignment of Disposition Codes and Response Rate Calculation	97
18. Survey non-Response	101
19. Presentation of Statistical Data	105
20. Revision of Statistical Data	111
21. Metadata Documentation and Quality Reporting	115
22. User Satisfaction Survey	119
23. Staff Satisfaction Survey	121

GLOSSARY

AO	Additive Outliers
CAPI	Computer-Assisted Personal Interviewing
CATI	Computer-Assisted Telephone Interviewing
CAWI	Computer Assisted Web Interviewing
COICOP	Classification of Individual Consumption by Purpose
CPA	Classification of Products by Activity
CPC	Central Product Classification
EC	European Commission
ESMS	Euro-SDMX Metadata Structure
ESQRS	ESS Standard for Quality Reports Structure
ESS	European Statistical System
ESS CoP	European Statistics Code of Practice
EU	European Union
FPC	Finite Population Correction
GDPR	General Data Protection Regulation
GSBPM	Generic Statistical Business Process Model
HICP	Harmonised Index of Consumer Prices
LS	Level Shifts
ISCED	International Standard Classification of Education
ISIC	International Standard Industrial Classification of All Economic Activities
ISCO	International Standard Classification of Occupations
ISO	International Organisation for Standardisation
IT	Information Technology
LAU	Local Administrative Units
LCI	Labour Cost Index
LFS	Labour Force Survey
KPI	Key Performance Indicators
MAR	Missing at Random
MCAR	Missing Completely at Random
ME	Margin of Error
MIG	Main Industrial Grouping
MSA	Malta Statistics Authority
NACE	Nomenclature statistique des Activités économiques dans la Communauté Européenne
NMAR	Not Missing at Random
NSO	National Statistics Office
NUTS	Nomenclature of Territorial Units for Statistics
PAPI	Pen-and-Paper Interviewing



GLOSSARY

PPS	Probability Proportion to Size
PRAM	Post Randomisation Method
PRB	Perceived Response Burden
PSU	Primary Sampling Units
RPI	Retail Price Index
SBR	Statistical Business Register
SBS	Structural Business Statistics
SDC	Statistical Disclosure Control
SDDS	Statistical Data Dissemination Standard
SDMX	Statistical Data and Metadata eXchange
SILC	Statistics on Income and Living Conditions
SIMS	Single Integrated Metadata Structure
SRS	Simple random sampling
SSU	Secondary Sampling Units
STBS	Short-term Business Statistics
TC	Transitory Changes
U	User Oriented
VPN	Virtual Private Network



01. Statistical Surveys

01.1 Scope and Purpose

In line with its commitment to implement appropriate statistical procedures, the NSO must ensure that every phase of a statistical survey is precise and correct. The success of a statistical survey depends on the accuracy of each stage, from planning to data collection, processing, analysis, and dissemination.

A statistical survey is a targeted enquiry conducted by a specific population to collect analytical information on a particular subject. This information is used for various purposes, including policymaking, assessment and comparability. Statistical surveys provide valuable insights, addressing potential problems or scientific investigations relating to the studied population. The target population may comprise individuals, private households, or other undertakings, such as enterprises, typically grouped for further analysis. Data may be collected from a census, a sample survey, or administrative sources in a statistical survey.

Statistical data can be collected, with the main methods being based on sampling, census and/or administrative data. *Sampling* involves selecting a subset (sample) from a larger population. The sample is selected to represent the entire population. By collecting and analysing sample data, statisticians can conclude the entire population without studying every individual.

A *Census* involves gathering data from every member of the population. A census aims to include every individual or element within the defined population. While a census can provide comprehensive and accurate information, it can also be costly and time-consuming compared to sampling.

Both sampling and census methods play crucial roles in statistical analysis. Sampling is often used when studying large populations for a more efficient data collection process, while a census is used when a complete view of the population is required. *Administrative Data* plays a vital role in official statistics, offering valuable information collected by government agencies and organisations during routine administrative processes, often used to complement or substitute other data collection methods.

01.2 Principles

A statistical survey includes planning, operations, and evaluation. The planning phase involves identifying users' needs, while the operational aspect focuses on the statistical process. This includes constructing the sampling frame, data collection and processing, analysis and publication. The evaluation phase entails documenting metadata and quality criteria and analysing user satisfaction surveys.

01.3 Recommendations

01.3.1 Designing and Implementing a Statistical Survey

- This phase must be implemented based on users' needs. Ensuring the collected data is relevant to the end user is essential when designing a statistical survey.
- Sufficient time must be invested in assessing end-user needs and the usability of the statistical data from the survey.
- When setting-up data collection, ensure that people, processes, and technology are in place to collect data and metadata.
- Resources must be adequately coordinated, ensuring sufficient human resources are available to meet objectives on time. Similarly, logistic resources must be organised, and adequate equipment must be provided so work is carried out efficiently.
- The mode of data collection must be determined based on the production schedules, costs, and the nature of the survey.
- A pre-defined timeline must be established (ideally using Gantt charts¹), with clear timeframes for each survey stage. This includes:
 - o Finalising the questionnaire;
 - o Organising fieldwork;
 - o Extracting the sample;
 - o Testing and allocating hardware;
 - o Printing of materials;
 - o Briefing interviewers;
 - o Monitoring data collection (including interviewer audits);
 - o Data entry;
 - o Coding;
 - o Benchmarking analysis;
 - o Publication.

The planning phase aims to achieve accurate statistical information. It determines data processing methods, estimation techniques and sampling design and is crucial to the survey's success.

01.3.2 Operational Phase

Sampling and Data Collection Procedures

- The sampling frame is developed during this phase, and the sample is selected. Adequate sampling frames must ensure coverage, completeness and timeliness.

Data Collection and Quality Control

- Quality controls must be implemented in the data collection stage to manage data quality and capture processes. Fieldwork operations must be monitored throughout the collection, with regular feedback provided to interviewers and field workers.
- Consistency checks must be conducted on key indicators throughout data collection. These indicators normally may include:
 - o Unit non-response rates (where a selected sampling unit such as an individual, household, or business fails to provide information);
 - o Item non-response rates (where respondents partially complete a survey but omit specific questions); and
 - o Proxy rates (where an individual responds on behalf of another person).

¹ A Gantt chart is a type of bar chart that visually represents a project schedule. It displays the start and finish dates of various elements of a project, such as tasks, activities, or phases, along a horizontal timeline.

Survey Performance and Data Processing

- Interviewer performance must be measured using Key Performance Indicators (KPIs) for self or interviewer-administered surveys. The KPIs should assess response and performance-based criteria, such as the following:
 - Characteristics of sampling units (e.g. sex, age and region of residence for individuals; number of members for dwellings, or number of employees for business sectors) and Paradata (e.g. response times, error logs), etc.
- Upon finalising data collection, the rate of over-coverage (see [5.2.1.2] Over Coverage) needs to be determined. This represents the proportion of units in the sampling frame that do not belong to the target population (i.e. they are out of scope). The rate can be calculated either unweighted or weighted, reflecting the overall frame or population rather than just the sample. The over-coverage rate is calculated as follows:

$$OCr_w = \frac{\sum_o w_j + (1 - \alpha) \sum_Q w_j}{\sum_o w_j + \sum_E w_j + \sum_Q w_j}$$

Where:

O is the set of out-of-scope units (over-coverage, resolved and not belonging to the target population)

E is the set of in-scope units (resolved units belonging to the target population of eligible units)

Q is the set of units of unknown eligibility

w_j is the weight of unit j

α is the estimated proportion of cases of unknown eligibility that are actually eligible. A conservative approach is normally considered and unless there is a strong reason to assume otherwise, all such units are considered eligible. In this case, α is set equal to 1.

- The input data from multiple sources must be merged using linkage variables (unique identifiers). The validity of linkage depends on the accurate matching of key variables, and the resulting linked records should be compared with expected data to assess reliability.
- The classification and coding of input data must comply with pre-determined standard international classifications or applicable national versions of these classifications.

Data Analysis and Dissemination

- Estimation and analysis methods must be carefully selected to ensure reliable results, considering the sampling design.
- Once available, the data must be benchmarked against (or compared to) previous cycles or other relevant sources to ensure consistency. Suitable disclosure control techniques (see [4] Statistical Confidentiality) must be implemented to uphold confidentiality rules and prevent breaches.
- To fulfil the main objective of data collection, survey dissemination products must be tailored to meet users' needs as closely as possible.

All necessary components for the timely release of statistical products must be prepared, minimising the time between production and release. Any pre-release embargo arrangements must align with internal dissemination policies. Statistical products must be actively promoted to maximise audience reach.

01.3.3 Evaluation Phase

Evaluation and Feedback after the Operational Phase

- After the operational phase, the National Statistics Office (NSO) or entity conducting the research must assess the overall success of the process and identify areas for potential improvement.

- The NSO must engage with experienced data users to gauge their satisfaction with the data provided.
- The evaluation must integrate staff recommendations with feedback from users.
- Process metadata must be reviewed to identify opportunities for improvement in each phase.

Formalising Documentation and Reporting

- Documentation of metadata, quality standards, and analysis of customer satisfaction surveys must be formalised.
- Outcomes from self-assessments and audits must be recorded in formal evaluation reports.

01.4 References

01. Statistics Finland. (2007). *Quality Guidelines for Official Statistics*. 2nd ed. Helsinki: StatisticsFinland.
02. United Nations Economic Commission for Europe. (2016). *Quality Indicators for the Generic Statistical Business Process Model [GSBPM]*. Geneva: UNECE.



02.

Accessing, Using and Managing the Use of Administrative Data

02.1 Scope and Purpose

Continuous advances in statistics are driving a growing demand for data. This growing demand creates financial burdens and often adds pressure on respondents. Respondents may become frustrated if they believe they have already provided the requested information for other administrative purposes. Technological advances enable statistical authorities to overcome the limitations of processing large datasets. Acquiring administrative data sources, which complement or sometimes can partially direct surveys, is a promising initiative.

Proper use of and access to administrative data supports the 2020 European Statistical System (ESS) vision on stakeholder cooperation. The ESS Code of Practice mandates the NSO with a clear legal mandate to collect, produce and disseminate European Statistics. The ESS Code of Practice mandates the NSO use administrative data for statistical purposes.

Statisticians use administrative data to:

- Update sample frames,
- Perform editing and imputation, and
- Conduct direct tabulation.

These data are also used for benchmarking validating collected data and estimates. Statisticians should remember that administrative data were not originally created for statistical purposes. Several limitations may arise when using administrative data, which statisticians must carefully address.

02.2 Principles

02.2.1 Legal Provisions

Statistical offices often rely on strong legal provisions to acquire administrative data for statistical purposes. The right to access records in Malta is defined by articles 38 and 39 of the Malta Statistics Authority Act (Chapter 422 of the Laws of Malta). Legal provisions guarantee statistical offices the right to access administrative records and enhance cooperation with data owners.

Article 39(1) (b) empowers the NSO to request and assist public entities in modifying how they collect data for statistical purposes. Article 39(2) requires public authorities planning activities with statistical implications to first consult with the NSO on the best data collection.

The legal framework regulating the use of administrative data sources should be enhanced by processes that address internal cultural barriers, where statisticians tend to favour conducting surveys instead of utilising data that is readily available from administrative sources.

02.2.2 Cooperation with Data Owners

Statistical authorities must improve cooperation with data owners to safeguard the timely provision and effective use of administrative data. Cooperation should go beyond acquisition rights. Statistical authorities and data owners must develop common strategies and standards to establish best practices for using administrative data. Providing administrative records to statistical offices in the required format may require data providers to invest time and resources. The importance of using their data for statistics must be clearly communicated and the benefits they may gain from sharing it.

02.2.3 Ensuring Trust and Security

Ensuring trust in official statistics is critical for fostering cooperation. The NSO implements safeguards to ensure the security and good governance of the entrusted data. Public confidence in data handling is essential. The NSO must minimise disclosure risks and ensure data are used solely for producing robust statistics. Data providers must be assured that the NSO requests access to administrative sources only when necessary for efficient statistical production.

For example, only authorised NSO officers who need the data to perform their tasks are granted access, ensuring confidentiality. The NSO must inform external users about how statistics are produced. The NSO must ensure that information about sources and methods is available to external users on its website.

02.2.4 Correctness, timeliness, and exhaustiveness

Statistical authorities rely on the quality of administrative data sources. Beyond technical accuracy, the correctness of data depends on the quality availability of metadata available for each source. Technical aspects include readability, population coverage, and the extent of missing data. Metadata aspects include clear definitions for population units and variables, comparability, and identification of unique keys.

Timeliness is another critical factor in using administrative sources. The prompt delivery of administrative is essential for publishing official statistics on time.

Exhaustiveness refers to compiling a comprehensive list of administrative data sources and maximising key national data sources. It also concerns the potential of these sources to cover the various target populations that statistical producers aim to study.

02.3 Recommendations

02.3.1 Legal Frameworks and Data Protection in Statistical Practices

- Legal frameworks must ensure access to administrative data sources and safeguard effective cooperation with data owners.
- Existing laws, if inadequate, must be revised to ensure the NSO is backed by a strong legal basis that accounts for differing legal stipulations from various data providers.
- The General Data Protection Regulation (GDPR) must be carefully considered. All administrative data obtained for statistical purposes must be treated as strictly confidential. Any statistical results must be treated as confidential, and statistical results should be produced without revealing details of any identifiable data subject.
- If there are doubts regarding data protection, advice must be sought from the NSO's Data Protection Officer. The Office manages data protection policies, promotes awareness, provides training, and ensures compliance.

02.3.2 Integrating New Data Sources into the NSO Statistical System

- The NSO should explore the potential of new data sources for integration into its statistical system. Whenever possible, administrative data should be used in place of direct surveys.
- The purpose and importance of each data source for the NSO must be clearly communicated. Acquired data must meet information demand and help reduce the response burden.
- Before integrating new data sources, the NSO must analyse their punctuality and delivery format.
- Fallback measures must be planned if the source fails to deliver data as agreed.

02.3.3 Quality Assessment and Compliance of Administrative Records for Statistical Use

- The NSO must assess the quality of administrative records to ensure they meet all prerequisites before use.
- Concept definitions in the data source must align with those of the particular statistical domain. If not fully compliant, methods must be explored to adjust these concepts to meet standard definitions.
- The units in the data source must belong to the target population.
- Unit and item non-response must be minimised, and the total number of empty cells must be determined (see [18] Survey non-Response).
- The data collection and inputting procedures at the source must be disclosed whenever possible.

02.3.4 Metadata Management and Data Source Requirements for Statistical Use

- Definitions of variables and population units must be recorded and stored in appropriate templates.
- The reference time of the data must be stated.
- Unique keys in the data must be identified. If unique identifiers are absent, combinations of variables that uniquely identify records must be established.
- Any limitations in using the data source for statistical production must be explained.
- Statisticians must request necessary metadata from data providers. This metadata enables statisticians to assess the fitness and suitability of administrative data, which was originally created for non-statistical purposes. It also allows both statisticians and users to utilise the data effectively.
- The NSO's metadata templates for describing administrative data attributes must be updated regularly.
- The template should include the following information:
 - o data source owner;
 - o frequency of data delivery;
 - o brief description of the data and its use;
 - o sector coverage;
 - o comparability of definitions with ESS requirements;
 - o statistical units;
 - o presence of unique identifiers;
 - o any known limitations.
- Make this metadata available internally to NSO staff.

02.3.5 Coordination and Management of Administrative Data at the NSO

- Directing the receipt of administrative data through key domain units or a dedicated coordination unit ensures that its use is restricted to a need-to-know basis while maximising its internal utilisation within the NSO.
- The NSO maintains a register of all administrative data used at the NSO and ensuring that metadata is compiled and regularly updated. It will also explore and identify new data sources that can be integrated into the national statistical system.

- The register must categorise each data source as follows:
 - Sensitive;
 - Restricted;
 - Internal;
 - Public.
 in line with stipulations of ISU-PRO-002 – Information Classification and Handling Procedure.
- The NSO will periodically evaluate administrative data sources every two to three years.
- Internal unique identifiers for persons, businesses and buildings need to be assigned to enable the linking of multiple data sources
- Employees must be informed about the optimal use of administrative data, its advantages, limitations, and the available data sources..
- Develop or enhance cooperation agreements with data providers to facilitate data sharing and ensure secure transmission.
- Cooperation agreements should follow the standardised NSO template where possible to ensure all relevant arrangements are included.

02.3.6 Proper Handling and Security of Administrative Data

- The proper handling of administrative data is crucial to:
 - preserving their confidentiality and integrity;
 - controlling access;
 - ensuring proper data retention and disposal.

02.3.7 Security policies and procedures need to be adhered to:

- ISU-POL-003 – Access Control Policy.
- ISU-POL-004 – Acceptable Use Policy.
- ISU-POL-005 – Information Security Incident Management Policy.
- ISU-POL-015 – Data Encryption Policy.
- ISU-POL-016 – Anonymisation and Pseudonymisation Policy.
- ISU-POL-019 – Personal Data Protection Policy.
- ISU-POL-021 – Data Retention Policy.
- ISU-PRO-002 – Information Classification and Handling Procedure.
- ISU-PRO-011 – Data Breach Response and Notification Procedure.

02.4 References

01. Daas. P et al. (2009). *Checklist for the Quality Evaluation of Administrative Data Sources*. The Hague, Netherlands: Statistics Netherlands.
02. Eurostat European Commission. (2017). *WP1 Access to administrative data*. Available: https://ec.europa.eu/eurostat/cros/content/wp1-access-administrative-data_en.
03. Royal Statistical Society. (2017). *Building public trust in official statistics*. Available: <https://www.statslife.org.uk/news/3575-building-public-trust-in-official-statistics>.
04. Statistics Canada. (2016). *Use of administrative data*. Available: <http://www.statcan.gc.ca/pub/12-539-x/2009001/administrative-administratives-eng.htm>



03.

Collaboration, Data Sharing, and Relationships with Data Providers

03.1 Scope and Purpose

The Malta Statistics Authority (MSA) Act of 2000, Chapter 422 of the Laws of Malta, and the European Statistics Code of Practice (ESS CoP) define the mandate and role of Official Statistics in Malta. The MSA Act establishes the MSA, which has the independent, impartial and accurate production of statistics, and the NSO, which produces these statistics. The ESS CoP also recognises the importance of professional independence and impartiality in statistical authorities.

The MSA Act empowers the NSO to collect data from persons, households, and enterprises. The Act also allows access to and use of administrative data for compiling official statistics, or if stipulated by ESS requirements. Only information that is necessary and relevant for such purposes is collected. The Act also permits access to and use of administrative data to compile official statistics.

To obtain data and release information, the NSO relies heavily on providing administrative data and its timely delivery. To maintain a strong, long-term relationship with administrative data providers and data subjects, with administrative data providers and data subjects, the NSO must ensure statistical confidentiality at all times and respect data providers. The NSO must also apply the principle of proportionality². Maintaining a sound relationship is crucial for the NSO to obtain good-quality data promptly.

03.2 Principles

When requesting data, the NSO must clearly specify the purpose of the request and how the data will be handled, in line with the principles and practices highlighted below.

- **Respect for data providers:** As the NSO collects mostly confidential information from individuals, households or enterprises consisting mainly of confidential information, it must approach respondents with respect. Respondents should be informed of the purpose of data collection, and confidentiality safeguards must be clearly communicated. The NSO should minimise the burden placed on them to maintain positive relationships (see [9] Response Burden).
- **Relationship between the NSO and Owners of the Administrative Data Sources:** The NSO establishes Memorandums of Understanding with the owners of the administrative data sources to enhance cooperation and to affirm that the data will be used solely for statistical purposes.

²By this principle, the reporting burden should be aligned with the needs of the users.

A good relationship between the NSO and administrative data owners can be built through regular meetings to discuss the continuous benefits of using these data sources. The NSO must communicate all security procedures to guarantee data confidentiality and integrity. Data owners should also be assured that only authorised employees will access the data, limited to those who need it for statistical compilation in their designated areas. It is also important to provide administrative data owners with value in return, such as customised outputs.

03.2.1 Processes to ensure that respondent information remains confidential:

- (a) **Oath of Secrecy:** All employees must take an oath of secrecy before accessing confidential data. By taking the oath, employees commit to not divulging any information encountered during their duties that could lead to the direct or indirect identification of individuals, households, or enterprises. This obligation remains in effect even after their employment with the NSO ends. Unauthorised disclosure of such information is a criminal offence under the MSA Act.
- (b) **Confidentiality and Data Protection:** The NSO checks all output shared with third parties or released to the public to ensure that individual data subjects cannot be identified. The NSO applies current best practices to ensure data confidentiality. Data protection complies with the ESS CoP, the General Data Protection Regulation, and the MSA Act. The MSA Act prohibits the disclosure of any information related to an individual, household or enterprise.
- (c) **Use of Administrative Data Sources:** The MSA Act authorises the NSO to use administrative data sources for compiling Official Statistics. Linking administrative data with survey data is encouraged since it lowers costs, reduces the burden on survey respondents, and enables new statistical analyses and outputs. The increased use of administrative data is reflected in Regulation (European Commission (EC)) 223/2009, as amended by Regulation (European Union (EU)) 2015/759, which facilitates greater access to administrative data sources for statistical purposes.

The NSO has the authority to assess the variables in administrative data sources and request any amendments to ensure their suitability for statistical purposes. Administrative data sources must also include sound metadata to use variables properly. By law, administrative data providers must seek and follow NSO's advice when creating new activities with statistical implications, ensuring the NSO can access and use data efficiently.

- (d) **Transparency:** To maintain transparency at all times, the NSO publishes metadata on its website related to the statistics it produces. These metadata provide details on the statistical methods employed to generate the outputs, the data sources utilized, and how administrative data sources are incorporated.
- (e) **Data Retention and Archiving:** Confidential data collected from individuals, households, or enterprises should be retained only as long as needed for statistical and research purposes. If the NSO requires this data for other purposes, it must be securely archived. Access to archived data requires prior approval. Such approval is granted strictly on a 'need-to-know' basis, allowing access only for specific tasks performed by statistical producers. When data are no longer needed, they are securely disposed of. The NSO retains identifiable records solely from the Census of Population and Housing. These primary Census records are transferred to the National Archives and remain closed to the public for 100 years from the Census date.
- (f) **Physical Security:** The NSO issues identification smart cards to all employees, allowing access. External visitors are identified before entering the premises; with their details recorded and accompanied by NSO officials during their visit.
- (g) **Information Security and Access Control:** Information security protects data during all stages: collection, processing, preparation for output, storage, and disposal. All NSO employees must adhere to the principles and guidelines ensuring confidentiality, integrity and availability of its assets. These assets include physical and electronic forms, such as paper questionnaires and electronic databases. The NSO must communicate information security principles are to be communicated to all employees, who must fully understand these guidelines.

The NSO enforces access control to safeguard information security. Employees access desktop and laptop computers using passwords they select, preventing unauthorised access. Data source owners may grant access only when sufficient justification is provided. Access rights are regularly reviewed.

Security controls need to be applied for the different categories of information received, processed, stored and sent by the NSO in-line with ISU-PRO-002 - Information Classification and Handling Procedure to ensure that all information is adequately protected. Security markings applied to documents indicate the required security level, with different handling approaches applied at each level. Microdata requires appropriate levels of security, with specific rules and procedures governing data access and processing.

The NSO's internal policy (ISU-POL-022: Cryptographic Key Management Policy) mandates that official portable media devices (e.g. USB flash drives) used for data transfer and storage are encrypted using the Advanced Encryption Standard (AES) encryption algorithm. Access to data stored on these flash drives is password-protected. Unauthorised individuals cannot access the data in the event of loss or theft.

Confidential documents must be securely locked away at the end of each business day, as NSO's policies require.

Firewalls protect the NSO's Information Technology (IT) system from external intrusion. External access to the NSO's IT systems is secured via a Virtual Private Network (VPN) connection.

- (h) **Processing administrative data:** A selected group of individuals should access administrative data sources on a 'need-to-know' basis, allowing them to perform their tasks within their specific domain. Only authorised staff should conduct record linkage with survey data. Limiting staff access to administrative data sources and confidentiality further safeguards and enhances data security. A system of pseudo-anonymisation, replacing original identification numbers with random sequential numbers, must be in place to prevent further linkage with other sources.
- (i) **Employee Training:** All NSO employees must receive confidentiality and information security training. This training ensures that NSO staff are aware of and knowledgeable about confidentiality and information security. Internal and external audits must be conducted to ensure all employees are safeguarding confidentiality and information security.

03.3 Recommendations

03.3.1 General recommendations

The NSO must:

- Collect only data necessary for compiling official statistics from individuals, households, enterprises or administrative sources.
- Maintain professional independence, impartiality and transparency at all times.
- Ensure statistical confidentiality is ensured at all times, preventing the disclosure of information about individuals, households, or enterprises.
- Review and update its security policies at least every 12 months.

03.3.2 Data providers and administrative data sources

The NSO must:

- Ensure that data from administrative sources is suitable for compiling official statistics. The NSO can request amendments to these variables for statistical purposes if needed.
- Use administrative data sources whenever possible to improve the quality of statistics and reduce the burden on data providers.
- Enhance its standing with public authorities. Administrative data providers must inform the NSO in advance when planning new activities with statistical implications, ensuring the NSO can provide input.

- Establish a Memorandum of Understanding with administrative data owners to create a cooperative framework for data exchange. Regular meetings should be held to strengthen these relationships.
- Respect data providers at all times, ensuring they are informed about the purpose of data collection and guarantee the confidentiality of their data.

03.3.3 Data retention and archiving

The NSO must:

- Retain confidential data only for as long as it is required.
- Ensure that then a domain no longer requires data access is removed. If confidential data is kept for future use, it must be securely archived.
- Ensure that confidential data that is no longer needed must be securely destroyed.

03.3.4 Security and processing of data

The NSO must:

- Ensure that access to the NSO's premises is restricted to employees, and visitors must always be accompanied.
- Enforce information security and access control to prevent unauthorised access to documents or data. Data access is granted strictly on a 'need-to-know' basis.
- Following the information security classification scheme, assign a security marking to sensitive data.
- Sensitive documents and news release content before publication must be appropriately labelled.
- Ensure that confidential printed documents are locked away at the end of each business day, and employees must lock their computers when leaving their desks.
- Receipt of administrative data should be made through key domain units or a dedicated coordination unit to minimise the number of employees who can access personal details.

03.3.5 Employee training

The NSO must:

- Provide mandatory and regular training to all employees on confidentiality and information security.
- Conduct internal audits and spot checks to ensure its policies are being followed.
- Ensure that disciplinary measures are in place in case any policies are violated.

03.4 References

01. Barrow, R. (2008). *Working together: how good relationships with providers can improve the quality of official statistics*. New Zealand: Reserve Bank of New Zealand.
02. Eurostat. (2011). *European Statistics Code of Practice*. Available: <http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15>. Last accessed 19th Sep 2017.
03. Groves, R M et al. (2017). *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Washington DC: The National Academies of Sciences, Engineering and Medicine.
04. OECD. (2017). *Good Statistical Practice Toolkit*. Available: <http://www.oecd.org/statistics/good-practice-toolkit/goodpractices/>. Last accessed 19th Sep 2017.
05. Statistics Canada. (2016). *Compendium of Management Practices for Statistical Organizations from Statistics Canada's International Statistical Fellowship Program confidentiality*. Available: <http://www.statcan.gc.ca/pub/11-634-x/2016001/section4/chap6-eng.htm>. Last accessed 19th Sep 2017.
06. Statistics New Zealand. (2017). *Privacy and confidentiality guidelines*. Available: http://m.stats.govt.nz/about_us/legisln-policies-protocols/privacy-confidentiality-gdlns.aspx. Last accessed 19th Sep 2017.



04. Statistical Confidentiality

04.1 Scope and Purpose

Under the ESS CoP, the NSO respects statistical confidentiality and ensures no identifiable information is disclosed. The office guarantees that the collected data is used exclusively for compiling statistics.

The ESS CoP principle on statistical confidentiality states that the privacy of data providers—such as individuals, households, enterprises, and others—must be protected. This commitment aligns with the General Data Protection Regulation (GDPR) and the MSA Act. All data collected by the NSO for statistical purposes must be treated as strictly confidential, and statistical results must be presented to prevent any data subject's identification. Protecting data confidentiality is vital to maintaining trust with data suppliers, ensuring complete and reliable data is available for statistical compilation.

Under the MSA Act, all employees must take an oath of confidentiality, obliging them not to disclose any information regarding individuals, households, or businesses during or after their employment with the NSO.

Before requesting information, data providers are informed that the information they provided will remain confidential and only be used for statistical purposes. Interviewers also explicitly inform respondents that all individual data will be kept confidential.

In recent years, collecting, processing, and accessing data have evolved, enabling more thorough statistical data analysis. The NSO provides Eurostat and accredited researchers with various statistical outputs and microdata sets. This facilitates further statistical analysis beyond the official statistics.

Statistical Disclosure Control (SDC) ensures that data sets are protected against the identification of data subjects so that results for the entire population can be derived while preventing the identification of individuals or entities. SDC methods must also ensure that reliable statistical data is available for decision-making and policy formulation. Such statistics are only provided in an anonymised form to prevent direct or indirect identification of data subjects.

04.2 Principles

04.2.1 Approaches to Statistical Disclosure Control

Different Statistical Disclosure Control (SDC) approaches are applied depending on the specific data set. Identifiable information can be inadvertently revealed through combinations of key variables when too few records exist, through matching with external files, or due to the overwhelming dominance of certain data contributors compared to others. When determining the methods for anonymising microdata and aggregated data (also known as *microdata*), the stipulations in policy ISU-POL-016 on Anonymisation and Pseudonymisation must be followed.

SDC protection is applied to both microdata and tabular data. Microdata refers to data sets where each variable for each respondent is recorded. Tabular data refers to information published in tables, where each cell represents several units contributing to its value. The primary objective of SDC is to ensure that no sensitive information about data providers is disclosed once the data set is made available to accredited third-party researchers.

04.2.2 Anonymisation of Microdata

When researchers receive microdata sets, these are fully anonymised, ensuring that individual respondents cannot be identified. In a microdata set, identifying variables are classified as direct or indirect. Direct identifiers include names, addresses, and ID card numbers, which can directly identify individuals. The first step in the anonymisation process is the removal or encryption of these direct identifiers.

Indirect identification occurs when a combination of key variables reveals a respondent's identity. Key variables, also known as 'implicit identifiers' or 'quasi-identifiers', do not disclose identities independently but can do so when combined.

Sensitive variables contain confidential information, such as income, that must not be disclosed. Although a key variable may not be sensitive, it could enable re-identification when combined with other variables. SDC methods are applied to both key and sensitive variables to prevent the identification of individuals.

04.2.3 Principle of k-anonymity

One method to ensure confidentiality is k-anonymity, where each combination of key variables must appear in at least k records in the sample. A general approach used by the NSO considers k to be equal to 3 so that at least 3 persons share the same identifiable information. If this is not satisfied, global recoding and local suppression methods are applied.

04.2.4 Global Recoding

Global recoding can be applied to both categorical (variables that represent categories or groups) and scale variables (numerical variables that measure quantities). Certain categories are grouped for categorical variables to create new categories with higher frequencies. For scale variables, recoding can be applied by grouping data into groups (e.g. quartiles or quintiles).

Top- and bottom-coding are also types of global recoding. In top coding, this upper limit replaces values above a certain upper limit. In bottom-coding, values below a lower limit are replaced by that lower limit.

04.2.5 Local Suppression

If k-anonymity is not achieved after global recoding, local suppression is applied. In this method, for combinations that do not meet k-anonymity, the value of one of the variables is set as missing. Some variables can be assigned a lower likelihood of suppression than others.

04.2.6 Perturbative Methods

Global recoding and local suppression can lead to information loss, reducing the analytical value of the data. To address this, perturbative methods may be applied.

Rather than suppressing values, perturbative methods alter (or 'perturb') the values from their true values, introducing uncertainty. Although k-anonymity might not be satisfied, the risk of correct re-identification is reduced because the perturbed values are not exact. This ensures that researchers cannot identify individuals, and external file matching will not reliably disclose identities.

While perturbative methods reduce information loss, they may create the false impression that no anonymisation techniques have been applied.

04.2.6.1 Micro-aggregation

This perturbative method is applied to scale variables, such as turnover or gross value added. Records are first partitioned into groups based on a set of variables. The individual values of the scale variable are then replaced with an aggregate value (typically the mean or median) calculated for each group. This method helps achieve k-anonymity while minimising information loss.

04.2.6.2 Adding Noise

This method is used on scale variables, where a random number (noise) is added or multiplied by the original value. This protects data from exact matching with external files. The noise is introduced in a way that preserves the mean and variance of the original data. The added noise is usually continuous, following a standard normal distribution with a mean of zero and a variance of 1. This characteristic generally aligns well with the continuous nature of the original data.

04.2.6.3 Data Swapping/Record Swapping

In data swapping, attribute values are exchanged between records. The key parameters are the swapping rate (the percentage of records to be swapped), the variables to be swapped, and the selection of records for swapping. The purpose is to create uncertainty for data users, making it unclear whether records correspond to real data elements. Not all attributes are swapped, and only true swaps are allowed, meaning the records have different pre- and post-swap values. The rates for swapping must consider the risk of re-identification, the uniqueness of the attributes, and the permissible level of data distortion. Typically, swapping rates fall between 1% and 5%. However, higher rates may be necessary for highly sensitive datasets.

04.2.6.4 Pseudonymisation

Pseudonymisation involves substituting personally identifiable information with pseudonyms to safeguard individuals' identities while enabling data analysis. This process replaces direct identifiers with unique identifiers, using methods such as unique-key encryption or a lookup table that assigns a distinct identifier to each direct identifier. This approach minimises the risk of privacy breaches while preserving the data's utility for statistical analysis.

04.2.6.5 Post Randomisation Method

Post Randomisation Method (PRAM) is a perturbative method applied to categorical data. It uses a transition matrix containing probabilities that values of a certain variable will change or remain the same. PRAM is a probabilistic method that yields different results when applied to the same data set. An example of PRAM is the Cell Key method, which is used on frequency count tables once they have been compiled. This approach entails introducing noise to specific cell values or frequency counts, making minor adjustments that do not drastically alter the overall distribution but add a degree of uncertainty. This technique helps protect sensitive information while still enabling effective data analysis. This method was used for the protection of the 2021 Census tables transmitted to Eurostat.

04.2.7 Testing for Outliers

Most data sets used for compiling official statistics contain records with outliers, namely scale variables having values that are abnormally lower or higher than the norm. These records carry a higher risk of identification compared to others. Outliers are managed through the process of top (or bottom) coding, which establishes a maximum (or minimum) limit for data values. Data points that exceed (or fall below) this limit are substituted with the threshold value. Top coding is commonly implemented at the 95th or 99th percentile; for example, values exceeding the 95th percentile may be capped at that specific percentile value, thereby reducing the impact of extreme high values. Bottom coding typically applies at the 1st or 5th percentile.

04.2.8 Anonymisation of Tabular Data

Most NSIs deliver outputs in tabular data, such as frequency tables, which count respondents, and magnitude tables, which

sum respondents' contributions towards a cell value for a scale variable. Even though tabular data are aggregated, they can still reveal information about respondents. Cells with a small number of contributors or cells where a few respondents account for a large percentage of the total may pose a disclosure risk.

04.2.9 Threshold Rule

A table cell is deemed sensitive if it is based on contributions from fewer than n respondents and includes sensitive variables. If only one or two sensitive variables are present, and the cell has fewer than n contributors, confidentiality flagging may not be applied, as it is unlikely to reveal additional information. Nonetheless, the choice not to flag is dependent on whether the risk of identification stays within acceptable limits, and these situations are generally assessed on a case-by-case basis, considering the sensitivity of the information involved.

If three or more sensitive variables are involved and the cell has fewer than n contributors, it must be flagged as confidential.

04.2.10 $p\%$ Rule

A cell is considered sensitive if the second-largest contributor can estimate the largest contributor's value within $p\%$ accuracy. This rule is commonly used to control the disclosure of business macro data. As a rule, p is equal to 10%. A cell is sensitive (i.e., disclosive) if the cell total minus the two largest contributors is less than $p\%$ (where $p=10\%$). The $p\%$ rule can be formulated as follows: $X - x_2 - x_1 < \frac{p}{100} \cdot x_1$, where X refers to the cell total, while x_1 and x_2 represent the two largest contributors respectively.

04.2.11 Secondary Cell Suppression

Other cells in the table must also be suppressed to protect sensitive cells fully. Otherwise, values could be calculated through subtraction from sub-totals. This process is known as secondary suppression. It is also required in linked tables, where suppressed values in one table must be suppressed in corresponding entries in other tables.

04.2.12 Use of Waivers

A waiver is a written agreement between the NSO and data providers in which the latter consents to the publication of their data. The respondent may be either an individual or an enterprise. Under this agreement, the NSO is not obliged to safeguard the respondent's anonymity, allowing the data to be published freely. These agreements must follow the standard templates designed by the NSO.

The concept of waivers is used when disseminating business statistics. To reduce confidentiality suppressions in business statistics, consideration is given to publicly available information from external or other administrative sources. The cell is considered safe for publication if data about the largest two contributors is publicly available.

Through the implementation of a waiver system the focus is on x_1 and x_2 with the objective of preventing the two largest contributors from estimating each other's values within a margin of $p\%$. If only the information for x_1 is publicly available in the administrative data, the cell cannot be published if x_1 can estimate x_2 within a margin of $p\%$. This is reflected by the formula: $X - x_2 - x_1 < \frac{p}{100} \cdot x_1$.

04.2.13 Passive Confidentiality

The principle of passive confidentiality applies to International Trade statistics. International Trade statistics often detail imports and exports by product type, trade partners, and values. For certain products or countries, only a small number of businesses may be involved, making the data more easily traceable to specific entities.

Passive confidentiality allows the dissemination of as much data as possible while protecting sensitive details only when necessary. Data on imports or exports is only suppressed if requested by the importers or exporters, who may feel their

interests would not be at risk if the data were published. This principle is recognised in both Intrastat and Extrastat legislation. In other domains (e.g., health or education), stricter confidentiality measures (like anonymization or active confidentiality) might be more appropriate due to the nature of personal or sensitive information involved.

Passive confidentiality is primarily applied in International Trade statistics, not only because of the bidirectional nature of trade data (e.g., an import in one country is an export in another), but due to specific characteristics of trade data and the balance between data utility and confidentiality protection.

04.3 Recommendations

04.3.1 General Recommendations

- Only information necessary for compiling statistics should be collected.
- Before release, all data must exclude variables that could lead to the direct or indirect identification of respondents. Appropriate SDC methods should be applied to prevent indirect identification.

04.3.2 Access to Microdata

- Access to microdata is only granted after a thorough evaluation of the application form by the NSO.
- Researchers must demonstrate sufficient knowledge of handling statistical data and show that no alternative sources exist for the research project's data requirements.
- If access to microdata is approved, a contract must be established between the researcher and the NSO, specifying the conditions for providing the microdata.
- Microdata will be fully anonymised according to the terms outlined in the contract. Before the anonymised microdata is provided, the researcher will be consulted on the key variables to minimise information loss. This will determine the level of suppression and global recoding needed.
- Researchers must not disclose data to third parties; all outputs must ensure that respondents remain unidentifiable.
- The NSO must review the researchers' outputs before dissemination to ensure that data confidentiality is preserved.

04.3.3 Release of Tabular Data

- The disclosure risk of statistical units in tabular data must be evaluated before release.
- Disclosure risk is high if tables contain three or more variables, with at least one being sensitive. Confidentiality flagging may not occur if less than three sensitive variables are present, depending on the assessment of identification risk on a case-by-case basis.
- If a table contains frequent sensitive cells or if sensitive cells are prominent in specific categories, the table structure must be adjusted for protection.
- The necessary SDC methods (such as global recoding and/or suppression) must be applied, considering the NSO's responsibility to protect the privacy and confidentiality of statistical units.

04.3.4 Written Consent

- In rare cases, identifying statistical information may be disclosed. Written consent from data providers must be obtained in such instances and should be available upon request.

04.3.5 Storage of Confidential Information

- Confidential information must always be stored in a secure environment, such as servers with appropriate controls to preserve confidentiality. Encryption must be used if the information needs to be removed from the secure environment.

04.3.6 Virtual Private Network (VPN)

- Employees accessing confidential information from outside NSO premises must use a VPN on official devices. This protects data both during transit and when stored. Devices should be locked when left unattended.

04.3.7 Exchange of Confidential Information

- When exchanging confidential information, whether within or outside the organisation, it must be ensured that the recipient is authorised to access the information. The 'need-to-know' principle should be followed, and only the minimum required information should be shared.
- If email is used as the exchange channel, confidential data must be encrypted or password-protected to prevent disclosure if the email reaches unintended recipients. The decryption key or password should be sent through a different channel or separate communication.

04.4 References

01. (2005). *Data Swapping*. Available: <https://stats.oecd.org/glossary/detail.asp?ID=6904>. Last accessed April 2017.
02. (2005). *Record Swapping*. Available: <https://stats.oecd.org/glossary/detail.asp?ID=6967>. Last accessed April 2017.
03. (2008). *VPN Security*. Hong Kong: The Government of the Hong Kong Special Administrative Region.
04. (2012). *Chapter 10: Email Security*. Available: http://www.intelligentedu.com/computer_security_for_everyone/10-email-security.html. Last accessed April 2017.
05. Benschop, T, Machingauta, C, Welch, M (2016). *Statistical Disclosure Control for Microdata: A Practical Guide*. The World Bank.
06. Castro, J (2009). *Statistical disclosure control in tabular data*. Barcelona: Universitat Politècnica de Catalunya: Dept. of Statistics and Operations Research.
07. Dalenius, T, P. Reiss, Steven (1978). *Data-swapping: A Technique for Disclosure Control*. Program in Computer Science and Division of Engineering, Brown University. Stockholm.
08. De Wolf, P.-P., Hundepool, A., Giessing, S., Salazar, J.-J., & Castro, J. (2014). *μ ARGUS User's Manual*. Netherlands: The Hague: Statistics.
09. De Wolf, P.-P., Hundepool, A., Giessing, S., Salazar, J.-J., & Castro, J. (2011). *τ ARGUS User's Manual*. Netherlands: The Hague: Statistics.
10. Domingo-Feffer, J, Oganian, A, Torra, V (2002). *Information-theoretic disclosure risk measures in statistical disclosure control of tabular data*. Spain: Universitat Rovira i Virgili, IIIA - CSIC.
11. Domingo-Ferrer, J, Torra, V (2004). *Privacy in Statistical Databases*. Barcelona: Springer.
12. E. Fienberg, S, McIntyre, J (2004). *Data Swapping: Variations on a Theme by Dalenius and Reiss*. Berlin: Springer.
13. Eurostat European Commission (2009). *1st meeting of the High-level Working Group on Statistical Confidentiality*. Luxembourg: Eurostat European Commission.
14. Eurostat European Commission (2016). *Recommendations for Confidentiality Management in Business Statistics in the ESS*. Luxembourg: Eurostat European Commission.
15. Ichim et. al. (2013). *Report on disclosure scenarios and corresponding disclosure risks for the EHIS*. Luxembourg: Eurostat European Commission.
16. M. Stine, K, A. Scholl, M. (2010). E-mail Security: an Overview of Threats and Safeguards. *Journal of AHIMA*. 81 (Issue 4).

17. *Principle 5: Statistical confidentiality*. Available: <http://ec.europa.eu/eurostat/web/quality/principle5>. Last accessed April 2017.
18. Sanil, A, Gomatam, S, F. Karr, A. (2003). NISS WebSwap: A Web Service for Data Swapping. *Journal of Statistical Software*. 8 (Issue 7).
19. T. Duncan, G, E. Fienberg, s (1998). *Obtaining Information while Preserving Privacy: A Markov Perturbation Method for Tabular Data*. Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
20. Tarantola, A. (2013). *VPNs: What They Do, How They Work, and Why You're Dumb for Not Using One*. Available: <https://gizmodo.com/5990192/vpns-what-they-do-how-they-work-and-why-youre-dumb-for-not-using-one>. Last accessed April 2017.
21. Templ, M, Meindl, B, Kowarik, A (2017). *Introduction to Statistical Disclosure Control (SDC)*. Vienna: International Household Survey Network



05. Sampling Frames and Coverage

05.1 Scope and Purpose

Samples need to be drawn from sampling frames to acquire information and estimates on specific target populations. Unlike censuses, surveys cover a portion of the population, and results are generalised to the target population. Sampling frames must be relevant, accurate, and up to date. They should fully cover the surveyed units, be complete, and contain correct information on each unit.

The target population includes all eligible units, while the sampling frame includes only those that can be reached. For example, only units with available telephone numbers can be selected in telephone surveys. However, units without a telephone number are still part of the target population.

Every sampling frame contains a complete list of eligible units from the target population. These may include individuals, households, or businesses. These frames are referred to as **list frames**. Additionally, samples can be drawn from **area frames**. These are specialised list frames where units are grouped by geographical area, and the survey population is located within these areas. Area frames are used when unavailable list frames or the survey is geographic (e.g. in censuses).

Frames often include auxiliary information about each unit, which is used to select the sample and conduct the survey. For instance, a register of individuals and households used as a sampling frame might include identification numbers, names, dates of birth, genders, telephone numbers, and residential addresses. Similarly, business surveys use the Statistical Business Register (SBR), which contains details such as company names, addresses, number of employees, and types of economic activities.

If all units in the target population are selected, completeness of information is achieved, and full coverage is obtained. However, **coverage errors** occur when discrepancies exist between the target population and the sampling frame. These errors include **under-coverage**, **over-coverage**, **misclassification**, and **duplication**.

05.2 Principles

Sampling frames are constructed and updated using both administrative and survey data. A unique identification number is essential for linking the frame to other data sources. The identity card number is used for Maltese nationals in frames containing individuals. Passport numbers, RefCom numbers, or administrative identifiers may be used for others. In business surveys, the unique identifier is the SBR ID.

When a register, such as the SBR or the Population and Housing Register, is used as the sampling frame, foreign units not part of the target population may be included. The target population can be identified using specific criteria such as:

- **Demographic characteristics:** For example, a survey might focus on individuals aged 16 and over. Business surveys may target only specific industries.
- **Geographical area:** Some surveys may be relevant only to certain parts of Malta, such as a survey of adult Gozitans on a topic specific to Gozo.
- **Group-specific criteria:** Units sharing demographic characteristics with the population may still be excluded if they belong to a different group. For example, individuals living in institutions (such as homes for the elderly) are excluded from surveys focused on private households.

05.2.1 Problems of Sampling Frames

Errors in coverage within a sampling frame can lead to incorrect inclusion or exclusion of units from the target population.

05.2.1.1 Under-coverage

Under-coverage occurs when some units in the target population are excluded from the sampling frame. This happens when new units are created after the last frame update. These units have no chance of being selected. Measures must be taken to estimate and reduce under-coverage. Failure to address this issue can lead to biased estimates and underestimating the target population.

05.2.1.2 Over-Coverage

Over-coverage occurs when units included in the sampling frame no longer form part of the target population. This happens when population units cease between frame updates and the survey fieldwork. These units may still be selected, leading to an overestimating population size and biased results. Over-coverage must be identified and corrected.

05.2.1.3 Duplication

Duplication occurs when the same unit appears more than once in the frame. For instance, a person may be listed twice with provisional and final identity card numbers. Duplication leads to overestimation of the population size and may bias estimates. Detecting duplication can be difficult and often happens during data collection.

05.2.1.4 Misclassification

Misclassification occurs when incorrect auxiliary information is assigned to units in the frame. For example, a person's date of birth may be incorrect, leading to errors in age-related data. Similarly, businesses may be misclassified by employment size, affecting the accuracy of surveys. Misclassification leads to coverage errors and biased estimates.

05.2.2 Characteristics of a good quality sampling frame

A frame is of good quality if it is relevant, accurate, up-to-date, and cost-effective. The closer it matches the target population, the smaller the difference between the survey and the target population. Relevance is also measured by how well the frame allows comparisons across surveys covering the same population.

Accuracy is determined by how complete the frame is. It must cover all eligible units in the target population and include well-established classifications. It is crucial to provide the necessary information that allows researchers to contact respondents. For example, telephone interviews cannot be conducted if a frame lacks telephone numbers. Units should be correctly classified, and errors, such as under-coverage, over-coverage, and duplication, must be addressed.

Cost-effectiveness depends on how much it costs to construct, maintain, and update the frame. A frame becomes more cost-effective if it is automatically updated through linkage with administrative data sources.

Timeliness refers to how current the frame is concerning the survey's reference period. Although there is always a time lag

between the last frame update and the survey period, measures should be taken to minimise this.

In practice, few sampling frames meet all these requirements. The goal is to select frames that meet these criteria as much as possible. This will impact data collection, processing, and the quality of the survey results. Accurate sampling frames help minimise bias, improve survey estimates, and reduce data collection costs.

Administrative data sources are crucial for maintaining sampling frames. Coverage problems arise when these sources are not available. The quality of administrative data directly affects the accuracy of frame updates. Statistical institutions must understand the characteristics and metadata of these sources to ensure they are suitable for updating sampling frames. Frames must use standardised concepts, definitions, and procedures that researchers and data users understand.

05.3 Recommendations

- Assess the coverage of any new sampling frame during the planning stages to ensure its quality and sustainability.
- Consider a multiple-frame approach if a single frame cannot cover the entire population. This may involve combining list frames, area frames, or multiple ones. However, a single frame is generally preferred. Auxiliary socio-economic data may be linked from administrative sources.
- Use the same sampling frame for the same target population to reduce the cost of running multiple frames where possible.
- In the case of area frames, check the reliability of maps.
- Ensure that users of a sampling frame understand its characteristics. For example, the SBR maintained by the NSO has a two-year time lag due to regulations under the European Framework, which require data for reference year N to be provided by year $N+2$. Time lags are common in administrative sources and are a disadvantage when tracking the constantly evolving business sector. Similarly, the population register will have time lags due to the need for constant updates.

05.3.1 Maintenance of the frame

- Keep the frame as up to date as possible concerning the survey's reference period.
- Update the frame promptly, including additions such as births and deaths and unit changes in or out of scope.
- Regularly update the information in the frame to reflect changes in the population.
- Benchmark the number of units in the frame based on key variables with other reliable auxiliary data sources to ensure it accurately reflects the actual characteristics.
- Set up procedures to detect and minimise under-coverage and over-coverage. This could involve comparing the frame with other administrative sources and updating it when necessary.
- Account for duplicates, out-of-scope units, births, deaths, and changes in unit characteristics. Make the necessary updates.
- If necessary and feasible, incorporate questions in the survey questionnaires to resolve coverage and classification issues. Sampled units should confirm or correct existing information in the frame.
- Use verified information collected during fieldwork to update the frame. This includes updating addresses, telephone numbers, and other basic unit characteristics.

05.4 References

01. Statistics Canada. (2003). *Survey Methods and Practice*. Canada.
02. Statistics Finland. (2007). *Quality Guidelines for Official Statistics*. 2nd ed. Helsinki: StatisticsFinland.
03. Turner, A. et al. (2003). *Sampling Frames and Master Samples*. New York: United Nations Statistics Division (UNSD).
04. Wallgren, A. et al. (2016). *Frames and Populations in a Register-based National Statistical System*. Sweden: BA Statistisksystem AB.



06. Questionnaire Design and Testing

06.1 Scope and Purpose

A questionnaire collects data on a specific topic through written questions. The success of a research project largely depends on the quality of the questionnaire. Poorly selected, incorrectly phrased, or poorly ordered questions can result in misleading responses. Once data collection begins, correcting errors caused by a flawed questionnaire becomes difficult.

A questionnaire must impartially address the study's information needs and avoid overwhelming respondents. Using clearly defined, factual questions should also minimise interpretation errors during data collection.

A systematic questionnaire design is crucial for data quality and minimising measurement errors. A well-constructed questionnaire is smooth, clear, and logical. In line with the requirements of the ESS CoP, it helps produce official statistics that 'accurately and reliably portray reality.'

Testing a questionnaire is essential before starting data collection. The testing process evaluates content, such as the interpretation of terms or concepts, and technical aspects, like the flow of questions through routing and filtering criteria. In addition to identifying inconsistencies, testing assesses how user-friendly the questionnaire is for interviewers and respondents.

06.2 Principles

06.2.1 Questionnaire Design

Designing a questionnaire involves several consecutive phases: developing a conceptual framework, writing and ordering questions, incorporating visually aiding design elements, and creating electronic versions. Translating the questionnaire into primary languages is a crucial step in its design.

Questions in a questionnaire may vary in form and purpose. There are two main types: screening questions, used to identify the eligible research population, and main questions related to the topic of interest. Any question type can be either open-ended or close-ended (pre-coded).

- **Open-Ended Questions:** Respondents answer open-ended questions without any influence from pre-defined codes. These questions seek spontaneous responses and avoid the bias of suggested answer lists. They are typically qualitative.
- **Close-ended Questions:** Close-ended (pre-coded) questions limit respondents' answers to a set of alternatives provided by the researcher. These answers follow a continuum (scale) or appear as unordered options (list). Pre-coded questions may allow for multiple selections, and in some cases, respondents can include options not listed.
- **Likert Scales:** Likert Scales are used when an item needs to be evaluated based on how well it differentiates between statistical units with high and low scores. These scales consist of several statements expressing a favourable or unfavourable position on a particular research area. Through this method, respondents provide answers along a continuum ranging

from a low negative response to a high positive one. Responses are assigned numerical values, which are summed to determine the overall attitude of respondents.

- **Scale Construction:** When constructing scales, a key decision is whether all points on the scale should be labelled with words (verbal scales) or numbers (numeric scales).

06.2.2 Questionnaire Testing

Two main methods are typically employed for testing questionnaires. These are:

- 01. Pre-field methods** involve test interviews that are not conducted in the same way as the actual data collection process. These methods may assess only specific parts of the questionnaire and are generally used in the early stages of testing. The aim is to establish a pattern of how respondents might proceed while completing the questionnaire.
- 02. Field methods** refer to testing under field conditions, where interviews are conducted exactly as they would be during the data collection. These methods are commonly used in pilot testing and generally involve larger sample sizes than pre-field methods. Field methods also allow for quantitative analysis and typically focus on the entire questionnaire rather than individual questions.

06.3 Recommendations

06.3.1 Questionnaire Design

06.3.1.1 General Recommendations

- The questionnaire design must primarily consider the statistical needs of data users. Only include questions that add value to the research outcomes.
- All questionnaires require informed consent, giving respondents sufficient information about what will be asked and how their participation will support the study's objectives. Informed consent must also ensure confidentiality.
- The order of questions is crucial. It is generally better to present general-topic questions before more specific ones. Starting with specific questions can bias respondents by revealing the focus of the study.
- Sensitive questions are naturally associated with a higher level of non-response. These questions, which may be perceived as intrusive, should not be asked at the beginning of the interview. If they are asked later, and the interview ends after these questions, useful data for analysis will have already been collected, making the respondent's earlier answers valuable.
- Classification questions, such as those concerning sex, age, district of residence, and employment status, are essential for categorising survey participants for cross-analysis. As these questions are typically unrelated to the main topic, they should be placed towards the end of the survey to maintain conversational flow.
- Open-ended questions should be minimised as they require more time to answer, which can lead to survey fatigue and lower completion rates. Responses to open-ended questions are also harder to quantify and analyse statistically, often producing a wide range of answers that can be challenging to interpret and categorise.
- Use similar questions from other surveys, especially those tested and proven effective when possible. This ensures comparability between surveys.
- Questions must be clear, concise, and well-organised to enhance the respondent's experience. They should be kept as short as possible for ease of understanding. If several questions begin with the same phrase, it is more efficient to introduce the phrase at the start of the text rather than repeating it throughout.
- Avoid questions with negative phrasing or those that lead to biased responses. Opinions should be solicited purely on the topic, without hinting at the desired outcome, through unnecessary adjectives or adverbs.
- Hypothetical questions about experiences that respondents may not have encountered should be avoided.
- Avoid double-barrelled questions that address multiple topics, as respondents may be confused if a single answer cannot cover both topics.
- Use skip patterns for question routing where applicable.
- Pre-defined categories must be mutually exclusive, meaning they should not overlap and must cover all possible

responses. These categories should reflect variations in the population as closely as possible.

- The questionnaire must be translated into a language that respondents understand. The translation should ensure the questions are functionally equivalent so respondents in different languages interpret them similarly. Definitions and examples should be provided for concepts that respondents may not be familiar with, and these notes should be placed close to the relevant question. Reference periods should be clearly specified in each question.
- Acronyms must be fully explained, although their use should be minimised. If an acronym is used repeatedly, it must only be explained at its first occurrence.
- Space for additional comments should be provided at the end of the questionnaire to allow respondents to offer further input.
- Sections, divisions, consistent fonts, and numbering should be used to enhance the visual presentation of the questionnaire.

06.3.1.2 Open-ended questions

Open-ended questions are appropriate in the following situations:

- When the researcher seeks information from specialists on a topic in which they have less expertise.
- When targeting a small group of people, where complex statistical analysis is unnecessary.
- After preliminary research, to identify the range of opinions and behaviours within the population.
- Respondents can record their concerns, additional information, or opinions at the end of the questionnaire.

06.3.1.3 Closed questions

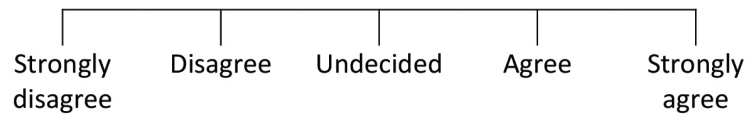
Closed questions are suitable in the following situations:

- For collecting demographic characteristics.
- When the researcher understands the topic and can provide an exhaustive list of possible answers.
- For telephone-administered questionnaires, there should be no more than six listed alternatives.
- In survey questionnaires, closed-ended questions may include an 'Other, please specify' option. This ensures that the response options are comprehensive, allowing respondents to choose from pre-defined answers while also providing the opportunity to specify responses that may not fit within the given choices.

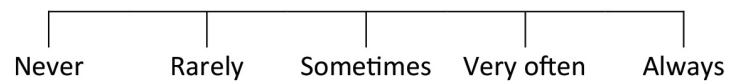
06.3.1.4 Likert scales

- The quality of Likert scales depends on how well they measure the attitude in question.
- The optimal number of alternatives should be determined by whether additional choices increase answer precision enough to justify the added complexity.
- Rating scales should have an odd number of points, with one neutral point (e.g., 'Neither satisfactory nor unsatisfactory') balanced by an equal number of positive and negative options.
- Once a scale format is selected, it should be used consistently throughout the questionnaire. While fewer choices make scales easier to use and more economical, they may lead respondents to feel that a desired option is missing. On the other hand, too many choices may leave respondents undecided between options.
- 'Don't know' and 'Don't remember' options should be used sparingly and not prompted by interviewers.
- Positive and negative options should be equally intense (e.g., 'Very difficult' vs 'Very easy').
- The endpoints of scales should be labelled. However, numerical scale values are more suitable for telephone interviews, as respondents are more likely to remember numbers. In self-completion and face-to-face surveys, verbal scales are more effective because they clarify the meaning of each point. Numeric scales only show equal divisions without explaining each point clearly.
- Five-point verbal scales for measured attributes should be constructed as follows:

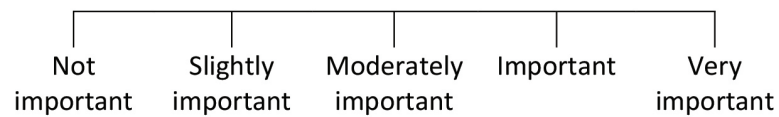
Agreement:



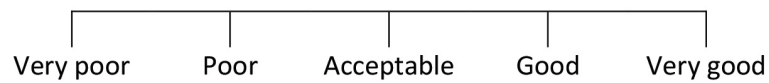
Frequency:



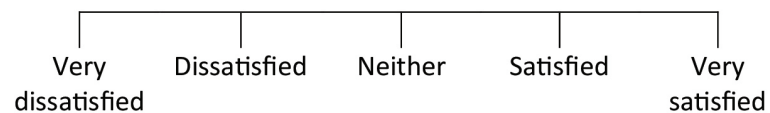
Importance:



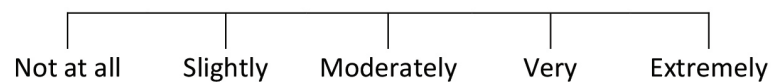
Quality:



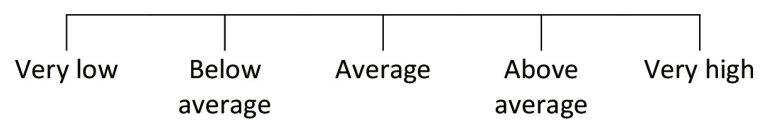
Satisfaction:



Relevance:



Value:



06.3.1.5 Electronic questionnaire design

Questionnaires designed for **Computer-Assisted Telephone Interviewing (CATI)** and **Computer-Assisted Personal Interviewing (CAPI)** have the same objectives as Pen-and-Paper Interviewing (PAPI). However, special considerations are necessary:

- The screen layout must enable interviewers to easily identify the text to be read and its location on the screen.
- Electronic questionnaires should allow for modification of previous responses.

06.3.2 Questionnaire testing

All types of questionnaires - whether self-completion or conducted by interviewers - must be tested in an organised manner before data collection begins. The results of these tests must be documented and considered when finalising the questionnaire. Testing is required in the following circumstances:

- When launching a new survey.
- When introducing new questions in existing surveys (e.g., ad-hoc modules).
- When modifying questions in existing surveys, even in the case of minor changes.
- When introducing a new data collection mode for existing surveys.
- When poor data quality is observed in existing questionnaires (e.g., inconsistencies, biases, or high non-response rates).

06.3.2.1 Field methods

- Field methods should assess the quality of the questionnaire by collecting feedback from both interviewers and respondents (debriefing). This feedback may be gathered through follow-up questionnaires after the field test interview. Additional insights can be gained from *paradata* (metadata about the data collection process, such as interview duration and interviewer response patterns).
- Systematic field testing should analyse response distributions, non-response rates, and editing and imputation rates.
- If software is used, it must be tested for inconsistencies with paper-based questionnaires. Any defects should be corrected before actual data collection begins.

06.4 References

01. Brancato, G. et al. (2006). *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System*. (European Commission Grant Agreement 200410300002). Available at: [Handbook_questionnaire_development_2006.pdf \(istat.it\)](#).
02. Lyberg, L. et al. (eds). (2012). *Survey Measurement and Process Quality*. New York: Wiley.
03. Scholtus, S., Van de Laar, R. and Willenborg, L. (2014). *Handbook on Methodology for Modern Business Statistics*. Available at: [246 \(europa.eu\)](#) (Accessed: 20 January 2022).



07.

Interviewer Recruitment, Training and Assignment

07.1 Scope and Purpose

The NSO is committed to implementing proper statistical procedures, ensuring that survey interviewers are trustworthy and capable of collecting accurate data. Interviewers play a vital role in surveys and must perform various tasks. Besides gathering information from respondents, they may need to locate addresses and provide valuable insights to update the sampling frame. Effective interviewing requires motivating respondents to participate, asking questions appropriately, and recording responses accurately.

Accurate statistical data depends on the interviewers' ability to collect precise information. A structured process for recruiting, training, and supporting interviewers is essential to maintain data quality.

Interviewer training should focus on developing the skills needed to conduct interviews, handle exceptional situations with respondents, and address any survey limitations. In addition to subject knowledge, interviewers must possess social competence, strong communication, and interpersonal skills.

07.2 Principles

07.2.1 Interviewer Bias

Bias arises when interviewers gather inaccurate information. This can occur for several reasons, such as mistakenly checking incorrect responses, misunderstanding concepts, or failing to record all details correctly. If interviewers do not ask the same question consistently, respondents may interpret it differently, leading to varying responses.

A significant concern is selective perception, where interviewers intentionally provide incorrect data to support their opinions, causing interviewer errors. Additionally, interviewers may falsify answers to follow certain skip patterns. This can involve skipping questions to finish the survey quickly or avoiding sensitive topics.

Sometimes, interviewers may need to probe questions to elicit more detailed responses when answers are incomplete or irrelevant. Probing involves using specific words or techniques to encourage a fuller reply. While careful probing is acceptable, improper probing can influence respondents and introduce bias.

07.2.2 Interviewer Recruitment, Training and Assignment

To reduce the risk of inaccurate data collection, interviewers must be carefully recruited, trained, and provided with timely assistance. When assigning interviewers to specific surveys, consideration should be given to the survey's topic and characteristics. The interviewer's familiarity with the subject must be evaluated to determine whether it will enhance data collection or risk introducing bias.

Key recruitment criteria include language proficiency, technical and computer skills, organisational ability, and availability. Previous interviewing experience is also a valuable asset.

Ongoing training is necessary to address unprofessional conduct and ensure interviewers stay updated with new developments. Regular, relevant training provides essential guidelines for efficient survey data collection.

Before any survey begins, interviewers must attend a briefing session. This session should explain the questionnaire clearly and provide a realistic overview. Key statistical concepts and critical aspects, such as questionnaire logic and eligibility criteria, should be highlighted. Interviewers need guidance on using the interviewing program correctly for surveys conducted electronically. The NSO should also designate a project leader to assist interviewers with any issues.

07.3 Recommendations

07.3.1 Interviewer Training

Areas to be covered during systematic interviewer training:

07.3.1.1 Professional Conduct

- Interviewers must be reminded of the professional conduct expected of them. As representatives of the NSO, they should be considerate towards respondents and, for face-to-face surveys, dress appropriately. They must carry their NSO identification card during fieldwork or clearly identify themselves during telephone surveys.

07.3.1.2 Introducing respondents to the survey

- Interviewers should understand the importance of a clear introduction, as it influences respondents' willingness to participate.
- After identifying themselves, interviewers must explain the survey's purpose and how participation contributes to valuable societal statistics. Confidentiality must be assured, and participation should be requested in a friendly manner.

07.3.1.3 Averting refusals and handling sensitive topics

- Training should focus on methods to minimise refusals. Interviewer attitude and personality significantly impact respondent participation. Interviewers must be trained to handle sensitive topics with empathy, as respondents are more comfortable with self-administered surveys on such issues.
- Interviewers should be informed that respondents are more likely to answer sensitive questions in self-administered surveys than interviewer-led ones. Therefore, interviewers must be trained to gather sensitive information while demonstrating empathy and maintaining respondent trust.
- For highly sensitive topics, such as experiences of violence, respondents may seek assistance. In these cases, interviewers must collect as much information as possible and provide details of available professional services when necessary.

07.3.1.4 Effective use of probing

- Interviewers must understand when probing is appropriate. Probing should only be used when respondents give incomplete, unclear, or irrelevant answers. Common probe questions include:
 - 'Could you give me an example?'
 - 'Could you explain that?' or
 - 'Could you be more specific?'
- Probing must always be neutral to avoid influencing responses. By repeating the full list, interviewers should clarify if a respondent chooses an option outside the pre-defined categories. If a respondent says, 'Do not know', a single probe should be used. Probing should stop once the respondent answers completely or shows signs of irritation.

07.3.1.5 Overview of the sampling design and sampling frame

- Interviewers should be briefly informed about the random sampling mechanisms and details of the sampling frame. This will help them provide consistent responses when respondents ask about their selection and the source of contact details, emphasising NSO's quality management practices.

07.3.1.6 Correct use of administrative forms

- The proper use of administrative forms, such as record sheets, non-contact forms, refusal forms, and daily progress sheets, should be explained during training. The correct use of these forms is important for the effective management of the survey.

07.3.1.7 Interviewer evaluation procedures

- Interviewers should be informed about the NSO's evaluation procedures during training. This will enhance their awareness of the importance of professional conduct.

07.3.1.8 Other training aspects

- Experienced interviewers also require regular training. They can share best practices with newly recruited interviewers, helping them improve their survey data collection skills.

Any relevant manuals or guidance notes should be distributed during these training sessions.

07.3.2 Other Considerations

- In large-scale surveys, more interviewers should be recruited than necessary. This allows flexibility in dismissing unsuitable candidates and selecting the most appropriate ones.
- Writing and reading skills must be assessed during recruitment. Additionally, poor eyesight may hinder interviewers from reading computer screens, so this should be considered.
- Interviewers should be paid fairly, considering the length and complexity of the questionnaire and any anticipated challenges in securing cooperation. Adequate compensation is essential to retain the most qualified interviewers.
- Once a payment strategy is set, the expected interviewer behaviour should be analysed. A balance between motivating interviewers and preventing rushed or incomplete interviews must be found.
- Interviewer safety must be prioritised. Surveys posing potential safety risks require special consideration.
- In panel studies, retaining the same interviewer for the same respondents across waves can be beneficial.
- Empathy and strong interpersonal skills are essential when assigning interviewers to surveys of a sensitive nature.
- Attention should be paid to 'veteran effects.' Experienced interviewers may struggle to adapt to new survey practices and developments in statistical data collection.

07.4 References

01. Alcser, K. *et al.* (2016). *Interviewer Recruitment, Selection, and Training*. Ann Arbor, MI: SurveyResearch Center, Institute for Social Research, University of Michigan. Retrieved February 15,2022 from <http://ccsg.isr.umich.edu/>.
02. Holm-Hansen, C. (2007). *Conducting Interviews*. Minnesota: Minnesota Office of Justice Programs



08. Methods of Data Collection

08.1 Scope and Purpose

Data can be collected through various modes, each affecting the entire statistical production process. The choice of data collection method requires careful consideration to ensure optimal results and should align closely with the production schedule.

Under the ESS CoP requirements, procedures must be established for planning quality statistical production processes. This includes systematic testing of questionnaires and making key decisions on appropriate statistical procedures, such as selecting the data collection mode before data collection begins.

08.2 Principles

Data can be collected from sampled units in various ways. Questionnaires may be either self-administered or interviewer-administered. Respondents complete self-administered questionnaires without assistance, while interviewer-administered questionnaires involve direct interaction between the interviewer and the respondent to collect the required information.

08.2.1 Paper-and-Pencil Interviewing (PAPI)

PAPI is one of the oldest and most common methods of data collection. It involves using paper questionnaires, either self-administered by respondents who receive them by post or completed through face-to-face interviews. In face-to-face interviews, the interviewer reads the questions from a printed questionnaire and records the respondent's answers.

08.2.2 Computer Assisted Personal Interviewing (CAPI)

CAPI involves collecting data using a computer or handheld device. Interviewers read questions to respondents and record the answers directly into the device. This eliminates the need for later data entry. The software includes built-in skip patterns, automatically directing the interviewer to the next relevant question based on the respondent's answers.

08.2.3 Computer Assisted Telephone Interviewing (CATI)

In CATI, the interviewer administers the questionnaire over the phone. The questions are displayed on a computer screen, and the interviewer records the respondent's answers. The system automatically advances to the next relevant question based on the recorded responses.

08.2.4 Computer Assisted Web Interviewing (CAWI)

In CAWI, respondents complete a questionnaire provided on a website. The flow of the questionnaire is customised based on their answers. As a self-administered method, respondents complete the online questionnaire without the interviewer's assistance. Another form of self-administered electronic questionnaires includes those sent as email attachments.

08.3 Recommendations

The choice of data collection method must align with the planned production schedule. Establishing deadlines for data availability helps set intermediate deadlines. The optimal method is chosen by assessing sample size, target audience, subject matter, questionnaire length, and associated costs.

Different data collection methods require varied testing procedures and collection durations. The adequacy and coverage of the sampling frame are critical in determining the appropriate method. For example, if the sampling frame does not include an up-to-date list of email addresses, CAWI cannot be utilised unless respondents are given (through postal mail or other means) personal credentials to log into a system and submit their responses.

In addition to the chosen methods, data can also be sourced from administrative records, which are pre-existing data collected by other entities. Using administrative data serves two purposes: reducing the number of questions (thus lowering the burden on respondents) or comparing the quality and accuracy of collected data with similar pre-existing information.

08.3.1 Timeframes, Planned Schedules, and Costs

- Self-administered and postal surveys typically involve lower costs.
- Face-to-face interviewing requires more resources and is the most expensive method.
- Postal and CAWI surveys may have unpredictable durations; reminders must be sent to non-respondents.
- For urgent data needs, electronic questionnaires are recommended. They save time in questionnaire vetting and data entry with instant validation checks and direct electronic data capture.
- CATI surveys are quicker than face-to-face or postal surveys. They are ideal when interim results are needed during the interview process.

08.3.2 Sample Considerations, Survey Topic, and Length of Questionnaire

- CATI suits small-scale surveys with fewer than (approximately) 30 questions and a small respondent sample. However, exceptions are occasionally made for larger surveys, such as EU-SILC and LFS, which are also conducted over the phone. This method is chosen when it is anticipated that the response rate will be enhanced due to a higher contact rate; and when interviewees themselves prefer to complete the survey over the phone.
- Ideally, long interviews should not be conducted over the phone. Telephone interviews should be limited to (approximately) 15 minutes. Exceptions to this are allowed based on the scenarios outlined in the previous point.
- Pre-coded questions with more than six options should not be asked over the phone.
- Sensitive questions should be self-administered, as respondents may feel their anonymity is compromised in interviewer-led surveys. In cases where only part of the questionnaire is self-administered, face-to-face interviews can be used for non-sensitive sections. Respondents can complete sensitive sections later, returning them in sealed envelopes. No reference numbers should be recorded in this case.
- Questionnaires with complex or technical questions often require an interviewer to clarify questions and answers. Complex questionnaires should be kept to a minimum, as a good questionnaire is smooth, clear, and logical.
- Self-administered electronic questionnaires should not be used if the target population is known to resist using technology.
- Large-scale surveys are well-suited for CAPI. CAPI enables pre-loading information from previous waves, which can be confirmed in subsequent collections. Exceptions are sometimes made if the response rates are otherwise improved or when respondents prefer phone interviews.
- Surveys targeting businesses differ from those for individuals or households due to differences in the target population.

Business surveys should be self-administered to allow enterprises the flexibility to respond at their convenience. Multiple individuals within a business may need to complete different sections of the questionnaire.

08.3.3 Testing Procedures

- Postal surveys typically have higher error rates because respondents cannot seek clarification on questions. They also experience higher non-response rates since there are no interviewers to follow up and encourage participation. Questionnaires for postal surveys must be thoroughly tested, focusing on how respondents will navigate and complete the form.
- Testing an electronic questionnaire takes longer than testing a paper version. Beyond wording, question order, and visual design, it is important to test the screen layout to ensure interviewers can quickly identify the text to be read. The correct routing of questions must also be verified.
- Testing should consider the time required for responses. A minimum response time must be maintained for telephone surveys.



09. Response Burden

09.1 Scope and Purpose

Response burden arises from the need to gather statistical information from finite populations. Individuals, households, and businesses must often participate in surveys, sometimes at regular intervals, such as monthly or quarterly. The demand for statistical data has grown significantly in recent years, driven by regulatory requirements or requests from government and private entities. Accurate statistics are essential for better decision-making.

Response burden refers to individuals, households, or businesses' efforts to complete a survey. It is typically measured by the time it takes to finish the survey and the associated costs. However, respondents' perception of the burden is also important. For businesses, the time spent on surveys could otherwise be used to increase productivity, pursue profitable activities, or enhance competitiveness.

As the ESS CoP outlines, the principle of proportionality requires that the reporting burden be proportionate to users' needs. The demand for statistics should not place excessive strain on respondents.

Statistical institutions must monitor the burden on respondents and set targets to reduce it over time. In addition to tracking the response burden, statistical offices should assess whether changes to questionnaires or data collection methods are necessary to ease this burden.

09.2 Principles

The following are key principles:

- **Actual Response Burden:** Refers to the time and/or money sampled respondents require to complete a questionnaire. This includes the effort involved in gathering the necessary information, filling in the questionnaire (for self-administered surveys), or providing the information during telephone or face-to-face interviews.
- **Perceived Response Burden:** Reflects respondents' views on how burdensome the questionnaire is. It measures the perceived time and effort required and the overall burden. The perceived effort, rather than the actual effort, often affects the responses' quality.

09.3 Recommendations

09.3.1 Frame and sample strategies

- The population and dwellings register, and the business register maintained by the NSO must be continuously updated. Regular updates add new units to the register, expanding the pool of eligible respondents.

- Updated information on units in the sampling frame ensures more accurate identification of the target population, reducing unnecessary contact with out-of-scope units.
- Selecting individuals or enterprises for survey participation must be reasonable, considering the frequency of participation in the same survey and the requirement to participate in different surveys.
- Measures to coordinate the sampling mechanism without introducing bias but reducing respondent burden should be explored. 'Negative coordination' helps ensure that a unit is not selected for multiple surveys within a short period, minimising overlap between consecutive samples.
- Thresholds should be set based on the mode of data collection, the effort required, and the number of contacts made with respondents for the same survey. These thresholds determine how long a respondent must wait before being selected again.
- While setting thresholds is relatively straightforward for individuals or households, it becomes more challenging for businesses. In a small country like Malta, the business population is often skewed. Large enterprises, dominant in terms of employees and turnover, are frequently contacted to ensure accurate representation of their sectors. To maintain their engagement, these enterprises must understand the significance of their participation in shaping business policies and decision-making.
- Any rotation method used to determine which units are included in the sample (or not to be sampled) needs to be aligned with the need for producing accurate statistics must align to produce accurate statistics. A balance must be struck between reducing respondent burden and achieving precise sample estimates that can be generalised to the entire population.

09.3.2 Collection Method, Timing and Assistance

- The ideal survey method should be selected with the type of respondents in mind.
- The chosen collection method should minimise handling costs and inconvenience for respondents. For example, providing pre-paid envelopes for postal surveys reduces the perceived response burden and eliminates financial costs for the respondent.
- Alternative participation methods should be offered to accommodate respondents' needs as much as possible.
- Electronic methods should be preferred for completing and returning questionnaires whenever possible. Easy-to-use survey websites with well-organised information can significantly reduce paperwork. When using electronic collection, alternative methods must be provided to accommodate respondents without internet access.
- Tight deadlines increase the burden on respondents and should be avoided when possible. Multiple survey deadlines for the same enterprise, sent by different units, should not overlap. Cross-unit coordination is essential; a centralised unit coordinating samples can facilitate this.
- Domain units must ensure that respondents' queries are handled efficiently and accurately. A lack of assistance increases the burden on respondents.

09.3.3 Questionnaire Design

- An efficient questionnaire design can significantly reduce the burden on respondents. If an evaluation reveals the need for redesign, attention must be paid to the structure and layout. Complex wording should be replaced with clearer, more understandable text, and necessary explanatory notes should be provided.
- The number of questions is directly linked to the perceived burden. Only essential questions for compiling statistics should be included and skip routines must be designed efficiently.
- Before completing a survey, respondents should receive clear information on its scope and importance for producing up-to-date statistics. Informed consent must be obtained for face-to-face or postal surveys, and telephone surveys should start with a thorough introduction outlining the survey's purpose. This reassures respondents that their participation is valuable for policy formulation and decision-making.
- Pre-filled information for periodic surveys, can reduce the burden. Respondents can verify or amend historical data rather than starting from scratch. However, this should be used cautiously to avoid perpetuating past errors or underreporting changes.

09.3.4 Administrative sources and data sharing

- Before conducting a survey, information should first be sourced from administrative records and registers. Administrative sources should be utilised only if they satisfy specific statistical requirements and maintain data quality, even after considering any data mining applied to such administrative data.
- Information sharing between functional units or other national authorities should be considered to prevent asking the same respondents identical questions multiple times.
- A coordinated system within the office is essential. New statistical needs can often be addressed by minor additions to existing questionnaires rather than launching a full-scale survey. Therefore, an evaluation of the details requested across various questionnaires is recommended.

09.3.5 Improve relationships with data suppliers

- Trust from individuals and businesses and interest in the survey is important for reducing the perceived response burden. Respondents must be assured of confidentiality, and interviewers should emphasise this when making contact.
- Trust is further strengthened by clearly communicating the statistical institution's identity, role, objectives, and information about the NSO's statistical products.
- Survey results can also be shared with participants to demonstrate the value of the collected data and foster a strong relationship with the statistical unit. These results can be available on the institution's website or through personalised and / or free reports to enterprises contributing to specific surveys.

09.3.6 Measuring Response Burden

- A statistical office can measure response burden (particularly among businesses) using a combination of direct feedback, process monitoring, and proxy indicators. One approach is to gather **direct feedback** through post-survey evaluations, where businesses can provide insights into how much time they spent completing the survey, the effort involved in retrieving necessary data, and their overall perceptions of the survey's complexity and usefulness. Additionally, conducting interviews or focus groups can provide qualitative insights into specific pain points, such as unclear instructions or burdensome data requirements. However, this will result in additional response burden on its own.
- **Paradata analysis** is another key method. By measuring the time taken to complete the survey, the Office can get a sense of the overall effort involved. Monitoring for breaks or interruptions during survey completion, as well as analysing navigation patterns (e.g., revisiting questions or using help resources), can indicate where businesses experience difficulty. Tracking how long businesses take to complete the survey and any delays in submission can also help gauge the burden.
- **Proxy indicators** can provide additional insights. Response rates can help identify trends: declining response rates may signal increasing burden or survey fatigue. Similarly, high item nonresponse rates can indicate that businesses find some questions difficult or too burdensome to answer. Monitoring the use of existing records or systems for reporting, such as payroll or sales reports, can highlight how often businesses need to retrieve or calculate information, which could add to the burden. Furthermore, tracking resubmissions or corrections can point to challenges in understanding or completing the survey correctly.
- To understand burden more quantitatively, a statistical office can estimate the **burden hours** involved. This can be done by multiplying the average time taken to complete the survey by the number of respondents or by calculating the financial cost of compliance, based on the average hourly wage of the staff involved.
- The office can also use **sampling and rotation strategies** to determine how often businesses are selected for multiple surveys. Being selected repeatedly for surveys in a short time frame can compound the burden. Similarly, using **administrative data integration** allows the office to assess whether data already collected from other sources, like tax filings or company registrations, can reduce the burden of collecting new information.

- **Business profiling** can help identify which sectors or business sizes face the greatest burden. For example, small businesses may find surveys more challenging due to limited resources, while certain sectors may experience more difficulty due to the complexity of their operations. Engaging with **business representatives** or industry associations can further gauge the overall sentiment about survey burden and highlight areas for improvement.
- The office can also track **burden metrics** over time, such as average time per respondent or burden hours by business sector, to evaluate the effectiveness of any burden-reduction measures and adjust strategies accordingly. Additionally, longitudinal monitoring can help determine if changes in survey design or approach are reducing the burden over time.
- By combining these methods, a statistical office can measure and address response burden comprehensively, ensuring that data collection processes are efficient while minimising the effort required from businesses.

09.4 References

01. Bavdaz, M. et al. (2015). Response Burden in Official Business Surveys: Measurement and Reduction Practices of National Statistical Institutes. *Journal of Official Statistics*. 31 (4), pg559-588.
02. Dale, T. et al. (2007). *Handbook for Monitoring and Evaluating Business Survey Response Burdens*. Luxembourg: European Commission, Eurostat.
03. Eurostat (2011). *European statistics code of practice for the national and community statistical authorities*. Luxembourg: Eurostat.
04. Giesen, D. (2012). *Exploring Causes and Effects of Perceived Response Burden*. Heerlen, Netherlands: Statistics Netherlands.
05. Hedlin, D. et al. (2005). *Developing Methods for Assessing Perceived Response Burden*. Research report, Stockholm: Statistics Sweden, Oslo: Statistics Norway, and London: Office for National Statistics.
06. Jones, J et al. (2007b), *Measuring and Monitoring Response Burden in Business Surveys*, Newport, Gwent, United Kingdom: UK Office for National Statistics.
07. Jones, J (2012). *Response Burden: Introductory Overview Lecture*. Newport, Gwent, United Kingdom: UK Office for National Statistics.
08. Lorca, D et al. (2011). *Sampling coordination of business surveys in the Spanish National Statistics Institute*. Madrid, Spain: Spanish National Statistics Institute.



10.

Consistent Application of Standards on Concepts, Definitions and Classifications

10.1 Scope and Purpose

Statistics are typically compiled from multiple observations, often combined with other statistical information. To ensure consistency, actual observations must follow harmonised concepts, definitions, and classifications. These elements are essential for integrating statistics smoothly into the broader statistical system. Standardised concepts and classifications are critical tools in official statistics, as they promote coherence across data sets.

In line with the ESS CoP on sound methodology, the NSO must consistently apply standard concepts, definitions, and classifications. These standards, set by the NSO, should comply with European or national legislation. Adjustments to concepts, definitions, or classifications used in surveys must also address user needs.

Concepts define the subject of inquiry, the statistical units from which data are collected, and the population under study. They describe the characteristics of specific observations. Classifications provide a systematic grouping of values that variables can take. These classifications describe various types of statistics, including economic, social, and regional data. Categories within classifications group or order measurements accordingly.

10.2 Principles

A classification serves as a tool, language, and model. It is essential for coding and presenting data in an organised way. A statistical classification groups or categorises variables in a survey or administrative file, aiding in the production of statistics. It consists of discrete categories, often arranged hierarchically, with a main group level and further subdivisions into lower levels. Each observation is assigned to only one category, and the classification must account for all possible data observations.

10.3 Recommendations

10.3.1 Use of Standard Concepts and Classifications

- Recognised international classifications, concepts, and definitions should be used wherever possible.
- Statistical concepts and classifications must be included in statistical releases and publications and clearly reported in metadata.
- Categories must be well-defined and supported by explanatory notes, indexes, rules, and correspondence tables. These tables highlight differences between current and earlier versions of classifications and differences between various classifications.
- Classification changes, particularly those that break time series, must be documented in statistical releases and publications.
- Differences in concepts and classifications across datasets for similar statistics must be clearly explained.
- When administrative data concepts and definitions differ from those required for statistical products, they should only

be used if they closely approximate the required standards. The comparability of these administrative definitions and classifications with standard statistical requirements must be documented in metadata.

- The residual group 'not elsewhere classified' should contain relatively few cases compared to other groups.
- When a dataset contains too few records for detailed classification, aggregation of groups should be considered.

10.3.2 National Adaptations of Concepts and Classifications

- National concepts and classifications should only account for specific national characteristics not adequately captured by international standards.
- National classifications must be linked to standard international classifications. National groupings may occur at lower levels, but higher classification levels should remain comparable with international standards.
- National classifications must ensure each activity is assigned to one category. Classification groups should be mutually exclusive and exhaustive.
- Items at the same hierarchical level should reflect the same level of detail, maintaining balance in classification structures.
- Structured codes must be used for classification categories, with longer codes representing more detailed categories.

10.3.3 Household Concepts

- A private household may consist of one or multiple persons who provide for their essential needs, such as food.
- The housekeeping concept allows for the possibility that the number of households may differ from the number of housing units. Under this concept, multiple households may reside in one unit.³
- The household-dwelling concept considers all individuals in one housing unit as part of the same household, equating the number of households with the number of housing units or dwellings.
- The appropriate household concept should align with the requirements of relevant methodological guidelines.

10.3.4 Use of International Standard Classifications

10.3.4.1 International Standard Classification of Economic Activities in the European Community

- Nomenclature statistique des Activités économiques dans la Communauté Européenne (NACE) Rev. 2 is the European classification of economic activities, corresponding to the International Standard Industrial Classification of All Economic Activities (ISIC) Rev. 4 at the international level. NACE is considered the European version of ISIC.
- NACE has four levels: Sections (1-digit), Divisions (2-digit), Groups (3-digit), and Classes (4-digit).
- The classification does not distinguish between types of ownership, legal operations, modes of operation, or market versus non-market activities.
- Ancillary activities, which support principal and secondary activities, do not count towards determining the activity code.
- To determine the principal activity, the top-down method is used. This involves identifying the section with the highest value-added, followed by the division, group, and class with the highest value-added within that section.
- NACE Rev. 2.1 is an updated version of the NACE classification that will begin to be implemented across various statistical domains starting in 2025. By 2028, all statistical domains will be using NACE Rev. 2.1 for coding.

10.3.4.2 International Standard Classification of Occupations ISCO)

- International Standard Classification of Occupations (ISCO) 2008 classifies jobs into distinct groups based on the tasks and duties performed.
- A person is linked to an occupation through their current main job.
- The primary focus for classifying jobs should be the responsibilities involved during the production stage.. For example, a baker who bakes and sells products should be classified as a baker, not a salesperson.

³ This concept was adhered to for purposes of Census of Population and Housing 2011. The sampling frame, maintained at the NSO, comprises a set of households as defined by the housekeeping concept.

- The main job is determined by the tasks and duties requiring the highest skill level. Skill levels are based on the knowledge required, tools and machinery used, materials worked on, and the goods or services produced.
- If multiple jobs require the same skill level, the main job is the one that consumes the most time.

10.3.4.3 Classification of Products by Activity (CPA)

- Classification of Products by Activity (CPA) is a system used to classify goods and services.
- CPA is the European equivalent of the Central Product Classification (CPC).
- The classification includes all products traded domestically or internationally or held in stocks. It also covers non-produced assets such as land, patents, and licences.
- CPA mirrors the structure of the classification of economic activities, with products classified based on their industrial origin.
- CPA has six hierarchical levels, similar to NACE Rev. 2.
- The first four digits of CPA codes align with NACE Rev. 2. The fifth and sixth digits represent further categories and sub-categories.

10.3.4.4 International Standard Classification of Education (ISCED)

- **International Standard Classification of Education (ISCED)** 2011 classifies education programmes and related qualifications by education levels and fields.
- In the ISCED framework, an education programme is a set of educational activities designed to achieve specific learning objectives or tasks.
- ISCED 2011 covers recognised formal and non-formal education programmes. It does not include unrecognised informal programmes or random learning.
- The classification of education programmes in ISCED is primarily based on their content. However, the programme's duration can be used as a proxy to determine the appropriate ISCED level.

10.3.4.5 Classification of Countries

- The International Standard for Organisation (ISO) 3166 country codes are essential for coding countries in nationality, country of birth, residence, and more statistics.
- The 3-digit numeric code is recommended for official statistics, while letter codes help identify areas easily.

10.3.5.6 Classification of Territorial Units

- The Nomenclature of Territorial Units for Statistics (NUTS) provides a uniform breakdown of territorial units for producing regional statistics.
- Local Administrative Units (LAUs) classify geographical areas at local levels and are compatible with NUTS. LAUs have two levels:
 - LAU 1 (formerly NUTS level 4) broad breakdown that exists in most countries.
 - LAU 2 (formerly NUTS level 5) includes municipalities or equivalent units.
- Malta's classification based on NUTS/LAU levels is as follows:
 - **NUTS Level 0:** Malta
 - **NUTS Level 1:** Malta
 - **NUTS Level 2:** Malta
 - **NUTS Level 3:**
 - Malta
 - Gozo and Comino
 - **LAU Level 1 (NUTS Level 4):**
 - Southern Harbour
 - Northern Harbour
 - South Eastern
 - Western

- Northern
- Gozo and Comino
- **LAU Level 2 (NUTS Level 5):**
 - 68 localities in total

10.4 References

01. Hoffmann, E. and Chamie, M. (1999). *Standard Statistical Classifications: Basic Principles*. New York: ILO and UNSD.
02. Statistics Canada. (2019). *Statistics Canada Quality Guidelines*. 6th ed. Ottawa: StatisticsCanada.
03. Statistics Finland (2007). *Quality Guidelines for Official Statistics*. 2nd ed. Helsinki: StatisticsFinland.



11. Statistical Data Editing and Imputation

11.1 Scope and Purpose

Missing and incorrect data are common and often unavoidable in surveys. Item non-response can occur for various reasons. Respondents may choose not to answer a question due to unwillingness or inability. In some cases, they may provide incorrect answers, even intentionally. Errors can also arise from interviewers who omit questions, misword them, or record answers inaccurately. Mistakes are not limited to data collection and can happen during data entry. Additionally, data may be reclassified as missing during analysis if responses are identified as clearly erroneous or inconsistent.

Addressing these issues requires thorough checking through statistical editing and imputation. The editing process identifies missing or incorrect data and checks for inconsistencies. Imputation then replaces these missing or erroneous data points with values that restore coherence and consistency. Although it is possible to leave values as missing and compensate for this during estimation (e.g., using weights), this becomes complex with many variables. Therefore, imputation is generally preferred to achieve a complete dataset.

Imputation corrects for bias caused by item non-response and erroneous data. Complete datasets simplify analysis using standard software and ensure consistency when analysing different subgroups.

11.2 Principles

11.2.1 Types of Missing Values

Understanding the mechanism behind missing data is the first step in addressing it. Missing values are typically classified into three categories:

11.2.1.1 Missing Completely at Random

Data are classified as missing completely at random (MCAR) when the probability of missing data points is unrelated to both the target variable and any auxiliary variables. In this case, the non-missing data can be considered a random subset of the data. For example, MCAR may occur if a survey question is mistakenly skipped during an interview.

11.2.1.2 Missing at Random

Data are classified as missing at random (MAR) when the probability of a missing data point is unrelated to the target variable but related to other auxiliary variables. For instance, elderly respondents might be more likely to forget past events than younger respondents. In this case, the non-response is related to the respondent's age but not the variable itself. Ignoring this mechanism may introduce bias, so it is important to account for it by adjusting for non-response within relevant sub-groups.

11.2.1.3 Not Missing at Random

Data are classified as not missing at random (NMAR) when the probability of a missing data point is related to the value of the target variable and possibly to other auxiliary variables. For example, high-income earners may be less willing to disclose their income than low-income earners, making missing income values NMAR. In such cases, inferences would be biased unless the non-response propensity is factored into the analysis.

11.2.2 Types of Errors

Errors in survey data can be either systematic or random. Differentiating between them is important for applying the correct detection and treatment methods.

11.2.2.1 Systematic Errors

Systematic errors occur consistently across respondents. Common causes include misprints in questionnaires, repeated misinterpretation of questions, or errors during coding, such as assigning an incorrect code to a description. Another example is a sign error, where a respondent forgets to report a negative sign for a loss in self-employment income.

The 'unit measure error' is another type of systematic error. This occurs when respondents report data using an incorrect unit of measurement. For example, if respondents are asked for financial figures in thousands of euros but provide amounts in euros instead.

Systematic errors create bias in results and must be identified and corrected. Since their causes are known, they are generally easier to detect and resolve than random errors.

11.2.2.1 Random Errors

Random errors occur without any systematic cause. They arise unpredictably throughout the survey process. An example would be entering an extra digit when reporting a numerical value. These errors may result from mistakes by respondents, interviewers, or during data entry.

11.2.3 Data Editing Methods

Detecting missing values is usually straightforward. However, it is important not to confuse structurally missing values with actual missing values. Structurally missing values occur when a respondent is not eligible to answer certain questions. For example, employment-related questions would not apply to children in a survey. Structurally missing values may also arise from the questionnaire's design, such as when a follow-up question is skipped if a respondent answers 'no' to a preceding question. It is also essential to distinguish between zeros and missing values to avoid bias in results.

Error identification, and to some extent, the detection of missing values, typically involves establishing a set of editing rules or 'edits' that define a range of possible values for each variable. This process requires insight and prior knowledge of the survey subject to ensure that all potential invalid responses and inconsistencies are covered. Edits can be classified into two types:

- **Hard (or fatal) edits:** These rules must always be satisfied. If violated, the value is automatically flagged as an error. For example, in a Likert-scale survey with values 1, 2, or 3, a response of 4 would be inadmissible.
- **Soft (or query) edits:** raise doubt about a value's validity but do not automatically classify it as an error. For example, if a survey collects income data, a soft edit might set an upper limit for plausible values. Any response exceeding this limit would be flagged for further investigation.

11.2.3.1 Interactive Editing

Interactive editing, or computer-assisted editing, applies edits during or after data entry, allowing errors to be corrected immediately. For data collected through computer-assisted interviewing, built-in validation rules can flag implausible values, which the interviewer can check with the respondent in real time. Other interactive editing methods include re-contacting respondents and comparing their responses with similar respondents' or previous survey data in panel

surveys. While interactive editing is effective, it is also time-consuming.

11.2.3.2 Selective Editing

Selective editing focuses on a subset of data, applying interactive editing only to cases within that subset. This method recognises that editing is costly and time-consuming and may place additional burdens on respondents. Not all errors significantly affect the final estimates, so selective editing prioritises critical errors that influence results. A comparison between the available data and expected values is made to identify significant errors, which may be derived from modelling, past surveys, or external sources. The difference between observed and expected values helps separate critical errors from non-critical ones.

11.2.3.3 Macro Editing

Macro editing examines data at an aggregate level, unlike the previous methods, which focus on individual data points (microdata). There are two macro editing approaches:

- **Aggregation method:** Final estimates are compared to benchmarks from previous surveys or external sources before publishing results. The individual records contributing to the estimate are reviewed if discrepancies are found.
- **Distribution method:** This approach identifies outliers - observations that deviate from the data's overall distribution. Tools like box plots, histograms, and scatter plots help visually detect potential issues. The distribution method can be used alongside the aggregation method to identify problematic records affecting aggregate results.

11.2.3.4 Automatic Editing

Automatic editing uses computer programs to detect and correct errors without human intervention. The program checks for violations of pre-defined edit rules, flags erroneous data, and sets them to miss. However, complications arise when several variables in a record are inconsistent, and the exact source of the error is unclear. Based on the Fellegi and Holt (1976) paradigm, a common solution is to modify the fewest data points necessary to satisfy all edits. This turns the problem of identifying the error into a mathematical optimisation challenge.

11.2.4 Imputation

Missing data arises from either item non-response or values identified as erroneous, which are then reclassified as missing during the editing stage. Imputation assigns values to replace these missing responses. Missing data introduces bias into the final estimates. Imputation helps reduce this bias. However, since non-response rarely follows a purely random (MCAR) pattern, imputation often needs to be done separately for carefully chosen subgroups of data.

Some imputation methods are deterministic, predicting a unique value for the missing data. This can lead to similar units being assigned the same imputed value. In contrast, stochastic imputation adds a random component to account for some of the variance among non-respondents.

11.2.5 Deductive Imputation

Deductive imputation is used when the missing value can be logically deduced from other survey values. This is common in financial surveys where totals must balance. For example, if one value is missing but the total is available, the missing value can be calculated. It is also used in longitudinal surveys when past data for the same respondent is available, and other responses show no change in circumstances.

Deductive imputation is simple and reliable, assuming the errors have been correctly identified. However, its application is limited because not all variables can be derived this way.

11.2.6 Model-based Imputation

Model-based imputation predicts missing values using a model built from auxiliary variables. Common methods include cell mean imputation and regression imputation.

11.2.6.1 Mean Imputation

In mean imputation, auxiliary variables related to the variable with missing data are selected. The data is then split into categories based on these variables. The variable's *mean* value is calculated for each category based on respondents who answered the question and used to replace the missing values. Careful selection of auxiliary variables is necessary, as mean imputation assumes the missing data mechanism is random within each category. This method assigns the same value to all missing cases in a particular combination of auxiliary variables, which may not capture the variance of non-respondents. A random value can be added to incorporate stochasticity based on the variable's distribution.

An alternative is to calculate the median instead of the mean. Outliers influence the mean, while the median is more robust for skewed distributions. However, the median is theoretically harder to work with. For symmetrical data, both mean and median yield similar results.

11.2.6.2 Regression Imputation

Regression imputation predicts missing values by fitting a regression model with the variable to be imputed as the dependent variable and using auxiliary variables as predictors. The values of the auxiliary variables must be available even for non-respondents. A stochastic element is added by including the error term in the predicted value.

Correlated auxiliary variables are identified during model fitting, improving the model's accuracy. While building regression models is time-consuming, especially when many variables need imputation, this method provides good predictions if the relationship between the dependent and auxiliary variables holds for both respondents and non-respondents. If not, the data may need to be split into combinations by auxiliary variables where the relationship is consistent.

Regression assumptions, such as independent error terms and constant variance, must be satisfied. Violations can affect the model's validity. In some cases, transforming the variables can address these issues. For categorical variables, logistic or log-linear models can be used for imputation.

11.2.7 Donor-based Imputation

In donor-based imputation, missing values are replaced by observed values from other respondents (donors). Variants of this method are described below.

11.2.7.1 Hot-deck Imputation

Hot-deck imputation selects donors from respondents in the same survey. The response from another participant with similar characteristics replaces the missing value. For example, if a 30-year-old unemployed male does not respond to a question, the donor will be selected based on similar traits. Donors can be chosen in several ways. Sequential hot-deck imputation selects the nearest previous respondent as the donor. Sorting the data ensures similarities among sequential units. However, if multiple units in sequence have missing data, the same donor will be used for all. In random hot-deck imputation, data is split into categories based on auxiliary variables, and a donor is randomly chosen from the same category of respondents. For reliable imputation, there must be a sufficient number of donors. Nearest-neighbour hot-deck imputation selects the donor with the smallest 'distance' to the non-respondent based on auxiliary variables. If more than one donor meets the criteria, one is selected randomly.

The advantage of hot-deck imputation is that all imputed values are real, observed data. Its success depends on the dataset's ability to match donors and non-respondents closely.

11.2.7.2 Cold-deck Imputation

Cold deck imputation is a method of handling missing data by replacing missing values with pre-specified or external values rather than deriving them dynamically from the dataset being analysed.

11.2.8 Multiple Imputation

The methods discussed so far are single imputation techniques, where a single imputed value replaces each missing value. However, uncertainty about the imputed value remains. Multiple imputation addresses this by imputing each missing value multiple times, creating several complete datasets. Each dataset is analysed separately, and the final estimates are derived by combining results across all datasets. An advantage of multiple imputation is that it allows for the estimation of variance introduced by the imputation process.

11.3 Recommendations

11.3.1 Preventing Missing or Erroneous Data

- Minimise missing or incorrect values by considering these issues during the questionnaire design stage.
- Incorporate automatic routing and in-built validations for computer-assisted questionnaires.
- Ensure the question wording is clear and easy to understand, with relevant instructions and definitions to avoid misinterpretation.
- Pay special attention to sensitive questions. For example, offering income brackets instead of specific values in income-related questions may improve response rates.
- A well-designed questionnaire, appropriate data collection method, pilot testing, and thorough interviewer training can reduce item non-response and erroneous data.
- Anticipate missing or incorrect data by including questions that provide useful auxiliary information, such as demographic data. This information aids analysis and facilitates imputation.

11.3.2 Early Phases of the Editing Process

- Distinguish between structurally missing values and non-response by assigning specific codes during data collection or entry (e.g., '-2' for not applicable, '-1' for missing). Avoid using blanks for missing values and ensure that a value of '0' is recognised as meaningful and not left blank.
- Explore the availability of relevant administrative data early in the editing process. Ensure both sources contain appropriate variables to enable micro-level linkage between the survey and the register. Assess the accessibility, timeliness, quality, and comparability of administrative data before using it as a tool in statistical editing or imputation.
- Panel surveys, where the same sampling units are surveyed over time, incorporate micro-level data from previous runs into the current survey's editing and imputation process. Verify data consistency both cross-sectionally and longitudinally. Values can be carried forward from the previous period while adjusting for inflation or percentage changes within relevant subgroups.

11.3.3 The Data Editing Process

- Begin editing at the micro level, followed by macro-level checks. The extent of micro-editing depends on available resources, such as time and budget.
- Excessive data editing can introduce as many errors as it corrects. To be effective, editing should be viewed as a learning tool for preventing errors in future surveys, not just as a method for correcting the current survey.
- Document all edits applied during statistical editing. This documentation informs users about data quality and provides valuable insights for future surveys. Clear documentation ensures traceability and enables impact assessment.

11.3.4 The Imputation Process

- While imputation reduces item non-response bias, it can introduce new biases. When selecting an imputation method, consider the nature of the variable (categorical, continuous, etc.) and the availability of relevant auxiliary variables.
- Many imputation methods involve splitting the data into categories containing similar units. These must be as homogeneous as possible and contain enough units to ensure reliable imputation.
- In donor-based imputation, avoid overusing the same donors across multiple non-respondents. Repeated use of the same donor can distort the characteristics of those donors in the final dataset.
- If a respondent has missing values for multiple variables, use the same donor for all variables to ensure consistency across imputed values.
- After separately imputing variables, even through different methods, conduct a final consistency check to ensure coherence.
- Flag all imputed values in the final dataset to enable users to assess the impact of non-response on the estimates.
- Retain the original variables in a separate file during imputation, ensuring the imputed values do not overwrite the original data. This allows for traceability.

11.3.5 The Imputation Process – Quality Indicators

- Carefully select the right imputation strategy by balancing timeliness (time required to produce the imputed dataset) with the quality of the imputation method.
- Key quality indicators include the item non-response rate and the imputation rate. Detailed documentation of the imputation process is essential. Users of the final dataset must know the extent of imputation to understand its potential impact on results. When data is collected through personal interviews, non-response and imputation rates are typically analysed by survey interviewer to detect potential interviewer bias.
- Estimates after imputation will differ from those observed without missing data. The non-response error represents the difference between these estimates.
- If no distinction is made between imputed and observed values, the resulting variance will not accurately reflect the true variance. The sampling variance under imputation and the variance caused by non-response and imputation should be measured.
- For totals and means, an important indicator is the contribution of imputed values to key estimates. A large contribution may indicate a high non-response bias or variance. This indicator is calculated as:

$$\text{Contribution of imputed values} = \frac{\text{Total weighted quantity for imputed values}}{\text{Total weighted quantity for all final values.}}$$

11.4 References

01. De Leeuw, E. D., Hox, J. and Huisman, M. (2003). 'Prevention and Treatment of Item Non-response'. *Journal of Official Statistics* 19 (2): 153-176.
02. De Waal, T., Pannekoek, J. and Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. New York: Wiley.
03. De Waal, T., Pannekoek, J. and Scholtus, S. (2012). 'The Editing of Statistical Data: Methods and Techniques for the Efficient Detection and Correction of Errors and Missing Values'. *Wiley Interdisciplinary Reviews: Computational Statistics* 4(2): 204-210.
04. Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
05. Kalton, G. and Kasprzyk, D. (1982). *Imputing for Missing Survey Responses*. Alexandria VA: American Statistical Association.
06. Kalton, G. and Kasprzyk, D. (1986). 'The Treatment of Missing Survey Data'. *Survey Methodology* 12 (1): 1-16.
07. Lohr, S. L. (2022). *Sampling: Design and Analysis*. 3rd ed. London: Chapman and Hall/CRC Press.
08. Luzi, O. et al. (2007). *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys*. (a Recommended Practices Manual jointly authored by ISTAT Italy, CBS The Netherlands, and SFSO Switzerland).
09. Rubin, D. B. (2004). *Multiple Imputation for Non-response in Surveys*. New York: Wiley Inter-science.
01. Scholtus, S., Van de Laar, R. and Willenborg, L. (2014). *Handbook on Methodology for Modern Business Statistics*. Available at: 246 (europa.eu). (Accessed: 4 March 2022).
10. United Nations Economic Commission for Europe (1994). *Statistical Data Editing Volume No.1: Methods and Techniques*. Geneva: UNECE.
11. United Nations Economic Commission for Europe (1997). *Statistical Data Editing Volume No.2: Methods and Techniques*. Geneva: UNECE.
12. United Nations Economic Commission for Europe (2006). *Statistical Data Editing Volume No.3: Impact on Data Quality*. Geneva: UNECE.





12. Statistical Estimation and Analysis

12.1 Scope and Purpose

Surveying every individual in a population, whether private citizens or businesses, is usually not feasible for a question of financial and other resources. Instead, a smaller group, known as a sample, is selected to represent the population. Since not all units are included, *sampling errors* may occur. This refers to the uncertainty about the accuracy of the sample estimates in representing the entire population. Although it's unlikely that only unusual cases are selected, some samples may over-represent certain population subsets, distorting the results.

Sampling errors can be minimised by using appropriate sampling methods. The right approach ensures that the selected sample provides accurate, representative information that can be applied to the whole population.

Before starting data collection, it is essential to decide on the sampling method and the desired level of detail for the analysis. These decisions provide crucial insights into determining the sample's accuracy and, consequently, its size. This also ensures that both the sample and analysis methods are suitable and compatible. Most importantly, the method of data collection should not influence the results.

12.2 Principles

12.2.1 Sampling Techniques

12.2.1.1 Simple random sampling

Simple random sampling (SRS) is the most straightforward technique, where each member of the population has an equal chance of being selected. In this method, sampling weights are equal for all units. SRS is most effective when no defining characteristics within the population could affect the data being gathered.

12.2.1.2 Stratified Sampling

In many cases, key variables within the population, such as age, gender, or location, influence how individuals respond. Stratified sampling accounts for these variables by dividing the population into **strata** (subgroups) based on combinations of significant variables. These variables are identified by statisticians, who must balance their expert judgment with avoiding assumptions that could skew results.

After dividing the population into strata, an independent random sample is drawn from each. Several factors can influence the determination of the number of units selected from each stratum, such as the desired margin of error, known response rates, response distributions, and variability within the strata.

12.2.1.3 Stratified random sampling (proportional allocation)

In its most basic form, units may be allocated equally across strata. The statistician determines the sample size n , ensuring that each stratum in the sample reflects the same proportion as its corresponding stratum in the population:

$$n_h = n (N_h / N)$$

Where:

N = sample size;

n = population size;

n_h = sample size of stratum h ;

N_h = population size of stratum h .

12.2.1.4 Stratified Random Sampling (Based on a fixed overall Margin of Error and Proportion)

This technique selects units independently based on the allowed margin of error for each sample statistic. The proportion p represents the expected outcome for a specific indicator. When the proportion is unknown, it is typically assumed to be 50%, as this produces the largest sample size allowing for the maximum margin of error to be expected.

$$n_h = \frac{N_h}{\left(\frac{(N_h - 1)(ME^2)}{p(1-p)(1.96^2)} \right) + 1}$$

Where:

ME denotes the allowed margin of error.

1.96 denotes the critical value that corresponds to a 95% confidence level in a standard normal distribution (Z-distribution) which is considered as a rule of thumb (see [12.2.5.2] Confidence Intervals).

12.2.1.5 Stratified random sampling (based on fixed relative margin of error)

This technique selects units independently based on a fixed relative margin of error determined by the *mean* of a specific variable of interest. Units are allocated disproportionately to ensure the desired precision for each variable.

$$n_h = \frac{N_h}{\left(\frac{(\mu^2)(ME^2)(N-1)}{(1.96^2)(\sigma^2)} \right) + 1}$$

Where:

μ = population mean in stratum h ;

σ = population variance in stratum h .

12.2.1.6 Stratified Random Sampling (Neyman Allocation)

This technique minimises the sample's variance by considering the variation and size of each stratum. Different sampling rates are applied to each stratum to reduce the variance in the mean and total estimated value of the variable of interest. The sample is distributed based on variability within the strata, not just population size. Strata with higher variance receive larger sample sizes.

$$n_h = n \left(\frac{N_h S_h}{\sum N_h S_h} \right)$$

Where:

S_h = standard deviation in stratum h .

12.2.1.7 Systematic Sampling

Systematic sampling is typically used for selecting samples of dwellings or households as it ensures that no two selected units are too similar (e.g. neighbours) while giving each unit an equal chance of selection. The sampling interval ($k = N/n$) is calculated), where N is the population size, and n is the sample size. The population is then ordered based on criteria relevant to the collected data (e.g. residential address, number of members having specific characteristics such as sex, age group, etc.). A random starting point between 1 and k is chosen, and every k^{th} unit after that is selected, continuing until n units are chosen.

12.2.1.8 Cut-off Sampling

Only cases above a certain threshold for a particular variable are sampled in cut-off sampling. The cumulative value for the selected cases must exceed an arbitrary cut-off point, usually represented as a percentage of the total value for that variable in the population. The cut-off must be large enough to ensure a good sample but not so large that it makes the survey overly expensive.

12.2.1.9 Cluster Sampling

Cluster sampling has three variants: one-stage, two-stage, and multi-stage.⁴ In all cases, the population is first divided into primary sampling units (PSUs), which are fast, cheap, or easy to sample. Examples of PSUs may include specific regions in a Malta, schools in different towns, etc. Some PSUs are selected, often using SRS.

- In **one-stage sampling**, all units (secondary sampling units or SSUs) within the selected PSUs are included.
- If all elements within selected PSUs are included, the second-stage sample is considered a cluster sample.
- In **two-stage sampling**, only a subset of SSUs is selected from each PSU using SRS.
- In **multi-stage sampling**, the SSUs are further divided into third- and possibly fourth-stage units, with random selections made at each level until the final units are chosen for the sample. A wide variety of sampling techniques may be used to select SSUs, including SRS, Stratified Sampling or Systematic sampling.

The number of units selected at each stage must ensure that the total sample size meets the target.

12.2.2 Design Effect and Sample Size

The chosen sampling method affects how estimators for population parameters are calculated⁵. The estimator for the population mean $\hat{\mu}$ is the sample mean $\bar{x}$ and is calculated as:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Where:

$x_i = i^{\text{th}}$ sample element.

⁴ In reality, one-stage and two-stage sampling are just like multi-stage but with the number of stages being 1 and 2 respectively. However, understanding one-stage and two-stage will make understanding multi-stage easier.

⁵ Real-life 'actual' values describing the whole population such as population mean and population variance.

Likewise, the population variance estimator $\hat{\sigma}^2$ is the sample variance s^2 and is found by:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

If a variable is binary, taking values of 0 or 1, the parameter of interest is usually the population proportion. The sample proportion p estimates this. The variance of this estimator is calculated as:

$$\hat{\sigma}^2 = p(1-p)$$

Where n is the sample size, and p is the sample proportion.

These formulae can be adjusted based on the sampling design, considering different strata, clusters, or other design features. In some cases, multiple sampling techniques are used together. A common approach is first to stratify the population, then apply another sampling method within each stratum. This makes variance estimation more complex, as the stratum estimators are calculated using the second sampling method and then incorporated into the overall variance equation for the first method.

The efficiency of a sampling plan is measured by comparing the variance from a specific sampling design to the variance from simple random sampling using the same sample size. This measure is called the *design effect (deff)*. Under a given sampling design, the design effect on the estimator $\hat{\theta}$ is calculated as:

$$Deff_{\hat{\theta}} = \frac{Var(\hat{\theta})}{Var(\hat{\theta})_{SRS}}$$

The design effect measures the precision gained or lost when using a more complex sampling design than simple random sampling. A $deff = 1$ indicates the same precision as simple random sampling; $deff < 1$ means greater precision, and $deff > 1$ implies less precision.

The design effect is also useful for determining sample size. First, estimate the sample size required for simple random sampling. Then, apply the estimated design effect from a similar survey to adjust this size for a more complex design. Multiplying the simple random sample size by the design effect provides the necessary sample size under the complex design.

The choice of sampling method and sample size are closely related. Deciding the appropriate sample size n requires balancing statistical accuracy and cost. Statistical requirements, such as limits on margin of error, variance, or confidence level, must be met. Estimates for the mean and variance from past surveys or pilot studies can help with these calculations.

Additionally, the sample size must comply with relevant regulations. After determining the appropriate sample size, assess whether the survey cost is within budget. If the cost is too high, either reduce the sample size incrementally while meeting statistical requirements or consider a different sampling method. If budget constraints make meeting the desired statistical requirements impossible, the requirements must be adjusted, or the financial plan (budget, expenses, etc.) must be revised.

12.2.3 Sampling for Audit Procedures

During data collection, it's essential to verify the accuracy of responses provided by interviewers. Checking every response is impractical due to time and cost constraints, so a sub-sample is selected for auditing.

To calculate this sample, select an equal proportion of each interviewer's responses, assuming all interviewers are equally likely to have errors. A (fixed or variable) proportion is then checked from each stratum. The larger the proportion, the more thorough the audit.

This sub-sample sampling method depends on the original sampling technique and how the units were assigned to interviewers. To ensure diversity, systematic sampling is commonly used.

12.2.4 Properties of Estimators

The primary goal of sample surveys is to derive point estimates for the parameters of interest. Estimators need to be unbiased, consistent, and efficient.

12.2.4.1 Unbiasedness

An estimator $\hat{\theta}$ is unbiased if it is centred around the population parameter θ . An estimator is considered unbiased when its expected value $E[\hat{\theta}]$ equals the estimated true parameter. The expected value is the weighted average of all possible sample estimates for the population parameter. The estimation bias of an estimator $\hat{\theta}$ is: $Bias[\hat{\theta}] = E[\hat{\theta}] - \theta$.

The estimator $\hat{\theta}$ is said to be an unbiased estimator for the estimator θ if $Bias[\hat{\theta}] = 0$.

12.2.4.2 Consistency

An estimator is consistent if the estimate converges to the true population parameter as the sample size increases. The estimate will equal the population value when the sample grows to include the entire population. Consistency is achieved when increasing the sample size reduces the standard error of the estimate.

12.2.4.3 Efficiency

Efficiency relates to the variance of an estimator. The smaller the variance, the more efficient the estimator. An efficient estimator produces estimates that deviate from the true population parameter as little as possible. This property is also known as precision.

12.2.5 Measures of Sampling Errors

12.2.5.1 Standard Errors

The variability of point estimates for population parameters helps quantify the quality of those estimates. A key measure of this variability is the **standard error**.

Different samples will produce different sample means when estimating the population mean from a sample. The standard error of the mean measures the variability between these sample means. A smaller standard error indicates more reliable data.

The formula to calculate the standard error of the mean is:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Where:

s = sample estimate for the standard deviation.

This equation can be modified to calculate the standard error for a proportion.

$$SE_p = \sqrt{\frac{p(1-p)}{n} \left(\frac{N-n}{N-1} \right)}$$

Where:

p = proportion of successes as estimated from the sample.

The term $\sqrt{\frac{N-n}{N-1}}$ is known as the **finite population correction (f.p.c.)**. The f.p.c. adjusts results from classic equations that assume an infinite population. As the population size decreases, this assumption becomes less accurate, making the f.p.c. more important.⁶

12.2.5.2 Confidence Intervals

While it is not possible to calculate the exact population parameter (e.g., the mean) from a sample, it may be estimated it using a **confidence interval**. A confidence interval provides a range of values that likely contain the true parameter. The width of the interval depends on the **confidence level**, that may range between 0% and 100% (typically set at 95% for consistency). A 0% confidence level gives a single value, while 100% includes every possible value for the variable. The confidence interval is defined as $estimate \pm ME_{\alpha}$.

In turn, $ME_{\alpha} = z_{\alpha}SE_{\bar{x}}$, where:

- $SE_{\bar{x}}$ is the standard error of the mean (defined in 12.2.5.1),
- α is the confidence level,
- Z_{α} is the z-score for the chosen confidence level.

Common z-scores for confidence levels are:

α	Z_{α}
90%	1.645
95%	1.96
99%	2.576

The confidence interval is centred at the sample mean $\bar{x}$. The full formula is:

$$CI_{\alpha} = [\bar{x} - z_{\alpha}SE_{\bar{x}}, \bar{x} + z_{\alpha}SE_{\bar{x}}]$$

When constructing a confidence interval for the sample sum S , the formula is similar to that of the mean but centred at the sample sum:

$$CI_{\alpha} = [S - z_{\alpha}SE_S, S + z_{\alpha}SE_S]$$

When estimating a proportion, the CI is centred at the sample proportion p . The formula for the confidence interval for a proportion is:

$$CI_{\alpha} = [p - z_{\alpha}SE_p, p + z_{\alpha}SE_p]$$

Where:

SE_p is the standard error of the sample proportion (defined in 12.2.5.1),

For small sample sizes (typically $n < 30$), a t-score is used instead of the z-score. The t-score is used when the population standard deviation is unknown. The underlying assumptions when using the t-score is that the data follows the normal distribution

⁶ Some textbooks frequently cite the f.p.c. as being: $\sqrt{\frac{N-n}{N}}$

The $\alpha\%$ confidence interval is designed so that if an infinite number of samples were taken, the $\alpha\%$ of those samples would have confidence intervals containing the true mean. This is similar to saying there is an $\alpha\%$ chance that the true mean lies within the $\alpha\%$ confidence interval. The same holds if we are interested in analysing a proportion.

12.2.5.2 Coefficient of Variation

The coefficient of variation (CV) measures relative dispersion by showing how large the standard deviation is compared to the mean. It is calculated as the ratio of the standard deviation to the mean, usually expressed as a percentage:

$$CV\% = 100 \frac{s}{\bar{x}}$$

The coefficient of variation can be estimated from the sample data using the same formula.

The coefficient of variation is useful because, as a ratio, it is independent of the unit of measurement and dimensionless. It is particularly effective when comparing data sets with different units or means. However, the coefficient becomes sensitive to small changes when the mean is near zero. It is only meaningful for values from a ratio scale with measurable intervals and a fixed zero value.

12.2.6 Variance Estimation

Variance estimation becomes complex when dealing with complex sample designs (such as cluster sampling) or non-linear estimators, such as ratios. Exact variance estimation may not be possible in these cases, so approximations are used. The main methods for variance estimation are 'formula-based methods' and 'replication methods.'

12.2.6.1 Formula-based Method with Linearisation

This method estimates variance for non-linear estimators, like ratios or regression coefficients, under standard sample designs. It is widely used for estimating variance because it simplifies calculations by approximating non-linear statistics with a linear function, making it computationally efficient. However, it is considered biased due to approximation errors and assumptions that may not hold for small or skewed samples. Its accuracy improves with larger sample sizes, as the approximations better reflect the true variance.

12.2.6.2 Replication Methods

Replication methods create random sub-samples from the original sample. Each sub-sample is used to estimate the parameter of interest, and the variability between these estimates is used to approximate the variance. These methods are effective for complex sample designs.

12.2.6.3 Balanced Repeated Replication

This method is applied in specific designs, such as cluster sampling with two final-stage units or stratified designs with two elements per stratum. Sub-samples are drawn so that only one unit from each cluster or stratum is selected per sub-sample, creating a balanced set of replications. Replicate weights adjust for omitted units, and variance is estimated from the sub-samples. Modifications are needed for clusters with more than two units, and odd numbers of units can cause bias, which is a limitation.

12.2.6.4 Jack-knife Method

The jack-knife method splits the sample into PSUs. A PSU could consist of a single observation, a cluster of units, or other groupings. Replicates are created by removing one PSU at a time, adjusting weights to compensate for the removal, and estimating the parameter of interest for each replicate. The variance between the estimates gives the variance of the original estimate. This flexible method can be applied to various sample designs, but some may require extensive computation.

12.2.6.5 Bootstrap Method

Using the same design, the bootstrap method treats the sample as the population and resamples from it. This process is repeated many times to create independent sub-samples, which are re-weighted to estimate the parameter of interest. The variance is calculated from the differences between the estimates. Although computationally demanding, this method can directly provide confidence intervals.

12.2.7 Use of Auxiliary Data in Post-Stratification

Pre-stratification is not always possible, often due to insufficient information about the population or budgetary and regulatory constraints. Post-stratification at the analysis stage can correct for imbalances, enabling better non-response adjustment and calibration if the sample is not representative of sub-groups. This is particularly important when over- or under-sampling results in distributions that do not match the population.

Post-stratification involves comparing the sample data to an auxiliary dataset (e.g. register or other survey data) that reflects the population distribution across strata. Adjustments are then made to align the sample with the population distribution, correcting potential biases.

12.3 Recommendations

12.3.1 Simple Random Sampling (SRS)

- SRS is appropriate when little prior information about the population is available.
- A complete list of all units in the target population is required to implement SRS.
- When the units are geographically dispersed, SRS may fail to capture population heterogeneity and be difficult or costly to implement.
- SRS is not appropriate for sampling small minority groups in a population.

12.3.3 Stratified Sampling

- A stratified sample can provide the same or greater precision than an SRS of the same size, especially when there is homogeneity within strata and heterogeneity between strata. It can be an excellent method for sampling small minority groups in a population.
- Stratified sampling can be more cost-effective, as it may require a smaller sample size while maintaining precision.
- Careful definition of strata is essential, and sufficient information must be available for all elements in the sampling frame. Sampling within each stratum ensures even sampling and eliminates the risk of selecting a cluster of subjects.
- Proportional allocation should be applied when statistical weights are not used.

12.3.4 Systematic Sampling

- Systematic sampling requires a complete list of all units in the target population.
- This method can produce biased samples if the population has a hidden periodic characteristic.

12.3.4.1 Cut-off Sampling

- Cut-off sampling is cost-effective and time-efficient, particularly in industries where large businesses dominate economic activity. This approach is used in the Short-Term Business Statistics domain which produces indices rather than absolute figures. The sampling method considers only the units with the highest turnover as part of the target population. Specifically, only those units whose cumulative percentage turnover within each stratum is at least 80% are included. The strata are defined based on economic activity, using the NACE classification.
- However, it increases the risk of bias, as it is not a probabilistic sampling design and it does not capture developments in smaller units.
- The accuracy of cut-off sampling depends on the reliability of the data used to set the cut-off threshold. It is not

suitable for volatile, incomplete, or unreliable data.

12.3.4.2 Cluster Sampling

- Cluster sampling is the most efficient design regarding time and cost, particularly for large geographical areas.
- It does not require a complete list of population elements, only lists at the group level. Individual lists are needed only for selected clusters.
- Precision is highest when there is heterogeneity within clusters. However, sampling errors may be relatively large in such sampling designs.

12.3.4.3 Documentation of the Original Sampling Design

- Estimation designs rely on the original sampling design, so proper documentation of this information is essential.
- Key details should include design weights, analysis weights, and the definition of strata (if applicable).

12.3.4.4 Sample Information During Fieldwork

- Once the sample is selected, data collectors must receive a list of the units to be surveyed.
- The list should contain only the essential details the data collectors require, ensuring clarity and brevity.
- Typically, this includes identification numbers (e.g., Household Key, ID card, VAT number), communication details for contacting the units, and the stratum each unit belongs to (if applicable).

12.2.4.5 Assessing Reliability of Point Estimates

- Sampling, rather than a full population survey, introduces a degree of error.
- Estimates from a sample rarely match the true population parameters exactly, and different samples will yield different estimates.
- The reliability of point estimates should be assessed by quantifying the standard error, coefficient of variation, and confidence interval.
- Where possible, the design effect should also be estimated.

12.2.4.6 Selecting a Variance Estimation Method

- Exact variance estimation formulae should be used whenever possible, though complex sample designs and estimators often require approximation methods.
- Both linearisation and replication methods usually provide consistent variance estimators.
- Due to its simplicity, the linearisation method is generally preferred for simpler designs and estimators. For more complex designs, replication methods may be more suitable.

12.4 References

01. Arsham, H. (n.d.). *Statistical Thinking for Decision Making: Revealing Facts from Figures*. (Last consulted: 2 April 2017).
02. Banning, R., Camstra, A. and Knottnerus, P. (2012). *Sampling Theory: Sampling Design and Estimation Methods*. The Hague: Statistics Netherlands.
03. Cochran, W. G. (1977). *Sampling Techniques*. New York: Wiley.
04. Eurostat (2002). *Monographs of Official Statistics: Variance Estimation Methods in the European Union*. Luxembourg: Office for Official Publications of the European Communities.
05. Lohr, S. L. (2022). *Sampling: Design and Analysis*. 3rd ed. London: Chapman and Hall/CRC Press.
06. Simon, J. L. (1997). *Resampling: The New Statistics*. 2nd ed. Available at: www.resample.com/intro-text-online/.

12.5 Other Sources

01. Online Statistics Education: A Multimedia Course of Study (<http://onlinestatbook.com/>). Project Leader: D. M. Lane, Rice University.
02. Properties of Estimators. BS2 Statistical Inference, Lecture 2 Michaelmas Term 2004. S.Lauritzen, University of Oxford, October 15, 2004. Available at: <http://www.stats.ox.ac.uk/~steffen/teaching/bs2siMT04/si2c.pdf>.
03. Standard Errors and Design Effects. ReStore @ National Centre for Research Methods. Available at: <http://www.restore.ac.uk/PEAS/errors.php>.
04. Sample Size and Design Effect. Prepared by G. Shackman (2001). Presented at Albany Chapter of American Statistical Association, March 24, 2001. Available at: [Sample Size and Design Effect | PDF | Standard Error | Sampling \(Statistics\) \(scribd.com\)](#).
05. Wolfram Mathworld. Available at: mathworld.wolfram.com.



13. Indices and Indicators

13.1 Scope and Purpose

An index is an indicator used in economic and business statistics to measure changes in a particular variable over time. Initially, indices were used to track price changes, but now they measure various indicators, including cost of living, industrial production, turnover, and employment.

Indices help measure changes that are not directly observable. By tracking changes in the values of variables that influence activities, indices provide insights into trends like inflation and deflation. Governments use indices to monitor inflation, assess the behaviour of certain sectors over a brief period, assess the cost of living, and adjust policies accordingly.

13.2 Principles

Types of Indices

There are three main types of indices:

- **Price Index:** Measures changes in prices over time. It is the ratio of the price of a group of commodities in the reference year compared to the base year.
- **Quantity Index:** Tracks changes in volumes of commodities, such as those related to industrial production.
- **Value Index:** Measures changes in monetary values of variables such as employment, turnover, hours worked, and wages.

13.2.1 Simple Price Indices

Simple indices show the change in activity for a single item. These indices, also called 'elementary indices,' reflect the price change of an individual item from the base period (representing a specific time period chosen for comparison, typically assigned a value of 100) to a specific reference period. Simple price indices do not account for the quantities of the item in each reference period. Therefore, they are unweighted indices. The main types of simple price indices are:

- **Dutot Price Index:** The ratio of the arithmetic *mean* of prices for several items across two reference periods.
- **Carli Price Index:** The arithmetic mean of price ratios for several items over two reference periods.
- **Jevons Price Index:** The geometric mean of price ratios for multiple items over two reference periods.

In short-term business statistics, the **Jevons method** is used for the Industrial Producer Price Index, while the **Dutot method** is applied to the Industrial Production Index and price statistics. These methods follow standard methodological guidelines.

13.2.2 Reference Periods

There are three key reference periods in index theory:

- **Index Reference Period:** The period in which the index is set to 100. This typically refers to a specific year and changes when the index is re-referenced.
- **Price Reference Period:** The period used for comparison, such as the previous month or quarter. Prices from this period are used as denominators in index calculations.
- **Weight Reference Period:** The period from which statistics are used to calculate the weights.

13.2.3 Laspeyres Index

The Laspeyres index is a common formula used to produce indices. For price indices, it measures changes in the cost of a fixed basket of goods and services over time. For an individual item i , let p_{i0} represent the price in the base period, p_{it} the price in the observation period, and q_{i0} the quantity in the base period. The Laspeyres formula is expressed as:

$$\frac{\sum_i p_{it} q_{i0}}{\sum_i p_{i0} q_{i0}} * 100$$

The numerator represents the total expenditure during the observation period, while the denominator represents the total expenditure during the base period. The formula calculates the index by summing the expenditure for each item.

To use this formula, one assumes that consumers purchase the same quantity of goods in both the base and observation periods. The suitability of this method depends on whether the quantities purchased in the base year remain constant and relevant over time. This index works best when quantities do not change significantly from the base period.

For practical application, the equation is transformed as follows:

$$\frac{\sum_i \frac{p_{it}}{p_{i0}} W_{i0}}{\sum_i W_{i0}} * 100$$

The variable $W_{i0} = p_{i0} q_{i0}$ represents the base period expenditure for each item, calculated as $p_{i0} \times q_{i0}$. This formula provides the weighted average of price ratios for each item, with the weight being the base period expenditure.

The Laspeyres formula can also be applied to other characteristics, such as employment and turnover. In these cases, turnover or employment values from the reference period are compared to those of the base period to calculate the index.

Other indices, such as the Paasche and Fisher indices, may also be used but rely on current period information (e.g., quantities). Since Laspeyres indices can be calculated before current period data is fully available, they are often preferred.

13.2.4 Retail Price Index (RPI) and Harmonised Index of Consumer Prices (HICP)

The Retail Price Index (RPI) and the Harmonised Index of Consumer Prices (HICP) are calculated using the Laspeyres formula. In the RPI, the weight of each item is derived from a fixed basket and updated whenever the basket itself is revised. RPI indices are calculated relative to base period prices. For HICP indices, the weights are updated annually, and the indices are calculated relative to the previous December.

13.2.5 Fixed Basket

The fixed basket of goods and services represents items most commonly consumed during a specific time, known as the base

period. This basket remains unchanged for several years, with revisions occurring periodically rather than annually. It serves as the basis for calculating the Retail Price Index, which ultimately determines a country's inflation rate.

Each item in the fixed basket is weighted according to expenditure figures from the base year. The basket is updated approximately every five years, following the Household Budget Survey (HBS). After the HBS is completed, the list of goods and services is reviewed, and weights are adjusted based on new expenditure data. Additional sources, like trade data and national accounts, may also inform expenditure updates, depending on the availability and reliability of HBS data for specific items.

13.2.6 Short-term Business Statistics (STBS)

Short-term business statistics (STBS) use three types of indices: price indices, quantity indices, and value indices. Growth rates are calculated to express changes in a variable over a specific period, either monthly or quarterly, allowing STBS to compute indices for current reference periods.

13.2.7 Price and Quantity Indices

Price and quantity indices are used to construct producer-price indices and production indices. To produce these indices, the price and quantity of products from sampled enterprises are collected monthly or quarterly. These values are then multiplied by their respective grossing-up weights to reflect the Business Register, representing the entire population. Aggregation weights, derived from the Structural Business Statistics (SBS) survey, are then applied to create indices at higher classification levels. These weights remain fixed for five years.

The growth rate is calculated by dividing the aggregated price or quantity value of the current period by that of the previous period. This ratio is then multiplied by the index of the previous period to obtain the current period's index. When new companies enter the sample, their contributions are excluded from the total during their first period to avoid distorting the data. They are included in the second period, ensuring their impact is accounted for in both the numerator and denominator of the growth rate calculation.

13.2.8 Value Indices

To calculate value indices, a similar method is applied to other indices. Values such as turnover, employment, wages, and salaries are collected from each company in the Short-Term Business Statistics (STBS) sample. These values are weighted using the grossing-up weight for each company, derived from the Business Register. The total for each sector is obtained by summing these weighted values. Sectors are typically defined using NACE or Main Industrial Grouping (MIG) classifications. The sector totals are then aggregated using weights from the SBS survey to calculate higher-level classification values.

13.2.9 Labour Cost Indices

Value indices also compute Labour Cost indices, which are calculated quarterly using the Laspeyres formula. Indices for wages, social security contributions, and overall labour costs are expressed relative to the base year. For instance, the wage index is computed by dividing the wage per hour worked in the reference period by the annual average hourly rate for the base period, then expressing the result as a percentage. The same approach applies to the calculation of other labour cost indices.

13.2.10 Chain Linking

Chain linking updates the weights more frequently, such as annually or every five years, rather than keeping them fixed for an extended period. The index is chain-linked each time the base year is updated, ensuring a comparable time series from one year to the next. This method provides results that better reflect real developments in the sector under review, as fixed weights over several years can lead to outdated results in some sectors. The annual overlap method is used for chain linking, following this formula:

$$CL_{t,q} = \frac{\sum W_{i,t-1} * I_{i,t,q}}{\sum W_{i,t-1} * I_{i,t-1}}$$

The variable $W_{i,t-1}$ represents the weights from the previous year, $I_{i,t,q}$ refers to the indices for the current month or quarter (q), and $I_{i,t-1}$ denotes the average indices from the previous year. The index for the reference period is then calculated by multiplying this chain link with the corresponding index from the previous reference year. The formula is expressed as follows:

$$I_{t,q} = CL_{t,q} * I_{t-1,q}$$

Where $I_{t-1,q}$ is the index for the same month or quarter of the previous year, and $I_{t,q}$ is the index for the current reference period.

This method calculates the HICP, where weights are updated annually, ensuring a comparable time series through chain linking. For HICP, the basket of consumer goods and services is updated according to the Classification of Individual Consumption by Purpose (COICOP). Chain linking is also applied in calculating GDP and its components and in STBS, where re-weighting occurs every five years for each sector. This ensures that the index using the old base year is linked and calibrated using the new base year. Chain linking is similarly used to produce Labour Cost indices.

13.3 Recommendations

- Only items directly related to the index should be included. For instance, non-dairy food items should be excluded from an index measuring dairy product consumption. Including the most representative items is also essential. If long-grain rice is the most consumed type, its weight must accurately reflect its importance in the index calculation. This consideration is addressed when selecting the SBS sample every five years, as the base year for STBS is updated. This sample ensures proper representation of products and NACEs for STBS transmission.
- To ensure indices accurately represent the economic situation, it is essential to include new companies or products that started operating after the base year but are significant contributors to the economy. Although these were not part of the original sample, their impact is considered in subsequent calculations. New companies or products are excluded from the growth rate during their first reporting period to maintain continuity in the time series. They are included from the second reporting period onward. Similarly, if a sampled company merges with a non-sampled company, the merged entity is treated as a new company, following the same approach as for new entrants.
- Adjustments are necessary when outliers appear in the data for a reference period. For instance, if a company has an unusually high performance for a specific period, grossing up its value may not provide a representative economic picture. In such cases, the grossing-up factor for that company is reduced to one, and the remaining factor is distributed among other companies in the sector.
- Replacements must be handled carefully. When an entity closes, a suitable replacement must be found. New entities are selected for the RPI and HICP indices based on similarity in price ranges, item availability, and turnover. Products must also be available regularly, and replacements should be similar to the original products.

13.4 References

01. Eurostat. (2016, September 27). *Short-term business statistics and (annual) chain linking*. Retrieved September 2017, 25, from Statistics Explained: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Short-term_business_statistics_and_\(annual\)_chain_linking#Annual_overlap_method](http://ec.europa.eu/eurostat/statistics-explained/index.php/Short-term_business_statistics_and_(annual)_chain_linking#Annual_overlap_method).
02. IMF. Basic Index Number Theory. In *Producer Price Index Manual*.
03. MBA- Lectures. (2010, June 20). *Laspeyres Price Index*. Retrieved September 25, 2017, from <http://mba-lectures.com/statistics/descriptive-statistics/543/laspeyres-price-index.html>

04. National Statistics Office. (2008). *Sources and methods for the compilation of the Retail Price Index and the Harmonised Index of Consumer Prices*. Malta.
05. Statistics Japan. (1996). *Q&A about the Consumer Price Index*. Retrieved September 25, 2017, from <http://www.stat.go.jp/english/data/cpi/1587.htm>
06. Statistics South Africa. (2013). *Note on Laspeyres, Paasche and Fisher Indices*



14.

Weight Construction and Adjustment for Non-Response

14.1 Scope and Purpose

Statistical samples often face imperfections, such as unequal selection probabilities, non-response, and non-coverage. These issues create differences between the sample and the reference population distributions.

To address these deficiencies and draw accurate conclusions, statistical weights are essential. Statistical weights aim to provide precise parameter estimates based on the chosen sampling method and estimation strategy. Weighting eliminates bias in estimates of totals, means, and proportions.

Sampling weights begin with constructing a design or base weight for each sampled unit. The design weight d_i is the reciprocal of the unit's selection probability p_i and corrects for unequal selection probabilities.

The design weight is then adjusted to counteract non-response bias, inaccuracies in sampling frames, and measurement errors. Further modifications may be applied to improve estimate efficiency and align with known population totals.

14.2 Principles

14.2.1 Design Weight in Sampling

The design weight forms the initial weighting strategy and accounts for sample selection. It directly reflects the sampling design, representing the probability of selection for each sampled unit. A unit's design weight (or sampling weight) equals its selection probability's inverse. Therefore, access to information on inclusion probabilities, even in complex multi-stage sampling designs, is crucial. Design weights inflate each sampled unit to reflect the total number of population units it represents. Summing these weights provides the total count of units in the sampling frame.

In general, sampling weights should be used to avoid biased estimates in survey data. However, weights may be unnecessary if the sampling design is self-weighting where each unit has the same selection probability and a constant factor can scale unweighted data to obtain accurate population estimates. Self-weighting samples are representative when each sampled unit reflects the population consistently. Parameters such as means and correlation coefficients, rather than population totals, are commonly estimated in self-weighting samples.

Weights can be rescaled by multiplying them by a constant, known as weight rescaling. This adjustment does not affect weighted means or proportions. In cases where estimates are needed only at the sample level, weights can be adjusted to sum to the sample total rather than the population total (see [14.2.6] Calibration for Reweighting).

14.2.2 Multi-Stage Sample Designs

Multi-stage sampling begins with an initial selection of elements from the population, known as Primary Sampling Units (PSUs). In the second stage, a sub-sample is drawn from each selected PSU, creating Secondary Sampling Units (SSUs). This process can continue with additional stages, selecting higher-stage samples as needed. If all elements within selected PSUs are included, the second-stage sample is considered a cluster sample.

For multi-stage designs, base weights must reflect selection probabilities at each stage. For example, in a two-stage design, where the i -th PSU is selected with probability p_i in the first stage, and the j -th SSU is selected from the PSU with probability $p_{j(i)}$ in the second stage, the overall selection probability for each element is given by $p_{ij} = p_i * p_{j(i)}$. The overall base weight is then the reciprocal of this combined selection probability.

14.2.3 Reweighting

Reweighting involves adjusting the original design weights by incorporating additional population and/or sample information. Variables considered for weighting adjustments should be predictive of survey outcomes. These variables form adjustment cells, with information typically available for all units in the sample. Main reweighting methods include post-stratification and ratio estimation.

14.2.4 Post-Stratification

Post-stratification is applied after all data collection is complete. It uses known population strata to adjust weighted survey counts, aligning them with reliable control totals for specific sub-groups, such as gender or age group. This method compensates for varying response rates across sub-groups, enhancing the precision of survey estimates. Post-stratification generates a scaling factor (or multiplier) that adjusts the weight within each cluster, ensuring that weighted sums for each cell match the population distribution.

14.2.5 Ratio Estimation

Ratio estimation adjusts sample weights using population totals for scale variables. Unlike post-stratification, ratio estimation is applied when external auxiliary information for the variable of interest is unavailable. The sample estimate is adjusted by calculating the ratio between the sample estimate of an auxiliary (known or measured) variable and its population total. This method works best when the auxiliary variable is strongly correlated with the unknown population parameter of interest.

For instance, if household food consumption is the variable of interest and external information is unavailable, you can adjust using the ratio of the sample estimate for household size to its population total. This adjustment works because food consumption and household size are closely correlated. Ratio estimation aligns the sample data with the population, enabling reliable inference to the entire population.

14.2.6 Calibration for Reweighting

Calibration is a commonly used reweighting method based on post-stratification. It adjusts the weighted sums of specific variables to match known or better-estimated population totals. This process reduces errors in population estimates derived from survey data, balancing bias and variance. Sampling weights are considered calibrated when estimates using these weights accurately reflect the population totals.

Calibration works by slightly adjusting sampling weights to achieve an optimal set of survey weights that sum to known population totals. The final weights are as close as possible to the original design weights but respect constraints to perfectly estimate the totals of all auxiliary variables in the calibration procedure. Calibration also helps reduce selection biases caused by unit non-response or frame errors.

A special case of calibration is *Raking*, also known as iterative proportional fitting or sample balancing. Raking starts with the initial sampling weights, which may already have been adjusted for unit non-response and non-coverage. These weights are then aligned with population control totals by cross-classifying categorical control variables and matching the total weights in each cell to the control total. The raking algorithm repeatedly adjusts weights across calibration variables until the weights converge and remain stable.

If weights are assigned based on clusters, such as age group and gender, the weighted sample will perfectly match population totals for these categories. However, these weights may not accurately reflect other important variables, like labour status, which are strongly related to the survey outcomes. As a result, the sample distribution by labour status may not align with the actual population totals.

To address this, the initial weights based on age group and gender are adjusted to ensure the weighted sample also reflects population totals for labour status. This adjustment is done using an iterative process until the weighted distributions closely match the population totals for all key variables.

The difference between initial design weights d_i and final calibrated weights w_i is measured by a distance function G . Obtaining weights close to the original design weights is an optimisation process that minimises G . Common distance functions include linear, logit, and truncated linear. Based on the chi-squared quadratic distance, the linear method converges quickly and reliably. The logit and truncated linear methods control weight ratios w_i/d_i (known as g -weights) with lower and upper limits, ensuring the weights remain within specified bounds.

14.2.7 Effect of Unequal Weights on Estimates

Reweighting ensures that final weighted data reflects sample design choices and improves estimate quality regarding precision and non-response bias. However, weight adjustments can lead to large weight variations, which may inflate sampling variances and produce unreasonable population means and total estimates.

Significant weight variations can result from probability proportion to size (PPS) sampling, response compensation, highly disproportional sample designs, and inaccurate data in multi-stage surveys. It is advisable to address the adverse effects of unequal weights during the design stage by increasing sample size to offset variance inflation from weight adjustments.

Trimming can be applied to limit weight variation and its impact on estimate variances, particularly for sub-groups. Trimming adjusts extremely low or high weights to fixed cut-off values. However, trimming may cause the total weights to no longer sum to the population size, which can be a drawback. In EU-SILC, the weights for each panel must fall within the range of [0.3 times the mean weight, 3 times the mean weight]. Trimming is performed whenever weights fall outside this range.

14.3 Recommendations

14.3.1 Weights for Non-Response Compensation

- Non-response occurs when not all desired information is obtained from sampled units. One way to address non-response is using a reserve sample, allowing replacements for non-responding units. Alternatively, a non-responding unit may be substituted with a nearby non-sample unit. However, this substitution increases the selection probabilities for substitutes and may introduce selection bias.
- Adjusting sample weights is the preferred method for managing non-response, known as non-response adjustment. Additionally, calibrating weights using external auxiliary information (if available) can further reduce non-response bias.
- Non-response adjustment involves modifying the design weights of responding units to account for eligible non-responding units. The non-response adjustment weight is calculated as the ratio of the gross sample size to the net sample size. The final weight is the product of the base and non-response adjustment weights. Researchers often view non-response treatment as a second sampling phase.

14.3.2 Weights for Duplicate Compensation in the Population Frame

- Some sampling units may appear multiple times in the population frame, increasing their probability of selection. To compensate, weights are assigned as the reciprocal of the adjusted selection probability for these units.
- For a sampled element i -th with an initial selection probability p_{i1} that appears k times in the population frame, with the remaining $k-1$ elements having probabilities $p_{i2}, \dots, p_{ik}$ for each duplicate, the adjusted probability of selection is given by: $p_i = 1 - (1 - p_{i1})(1 - p_{i2}) \dots (1 - p_{ik})$. The weight assigned to the sampled unit is then the reciprocal of this adjusted selection probability, calculated as $1/p_i$.

14.3.3 Calibration Weighting and Selection of Raking Variables

14.3.3.1 Selecting Raking Variables

- **Demographics like gender and age:** These are logical choices for ranking variables. Population totals for these variables are usually available and often have low non-response rates.
- **Consistent Population Totals:** A fixed total must be established if population totals for raking variables differ due to being sourced from various populations. Variables that do not sum to this fixed total need ratio adjustments.
- **Measurement Consistency:** Raking variables must be measured consistently across the survey sample and known population totals.
- **Variable Relevance:** Choose variables strongly related to key survey outcomes or those associated with non-response or non-coverage.

14.3.3.2 Ensuring Effective Raking

- **Convergence Factors:** The number of variables, their categories, sample cases in each category, and the discrepancy between the population total and the initial weighted sample total affect convergence. Set an upper limit on iterations to prevent non-convergence. Many iterations may indicate poorly behaved raking functions, resulting in problematic weights that inflate variances and produce unstable estimates.
- **Category Adjustment:** If a sample lacks cases in any category of a raking variable, revise categories or control totals (e.g., by combining categories). If any raking variable or control total category contains fewer than 5% of sample cases, collapse categories to avoid extreme weight disparities.
- **Handling Item Non-Response:** Impute missing values in raking variables before starting the process. If imputation is not feasible, adjust control totals to account for missing cases. Subtract the percentage of missing sample cases from the control total, then reallocate the remaining totals to maintain balance. This adjustment ensures that the weighted sample distribution reflects the same percentage of missing cases.

14.3.3.3 Selecting a Calibration Method

- Although all calibration methods are equivalent on average, they can produce different results for specific samples. The choice of method typically relies on empirical criteria.
- The **linear method** may result in negative or excessively large weights. To manage this, set constraints to keep weights within certain limits.
- **Bounded methods** (like logit or truncated linear) are generally preferred as they reduce the occurrence of extreme weights.
- The **limits on weight ratios** should not be arbitrary. Start with a small interval between limits and gradually enlarge it until a solution is found.
- Testing all methods to evaluate their impact on the weight distribution is advisable.

14.3.4 Weighting for Panel Surveys:

14.3.4.1 Adjusting for Panel Attrition

- In longitudinal surveys, attrition occurs when respondents drop out over time, reducing the initial sample size and potentially introducing bias. Attrition bias is more significant when drop-outs share specific characteristics linked to survey outcomes.

- Non-response adjustments for later waves of a panel survey can be made using data from previous waves. Responses from earlier waves provide detailed information that helps classify both respondents and non-respondents in subsequent waves. This data allows for more effective non-response adjustments compared to those made during the first wave.
- When adjusting weights for later waves of a panel survey, the process often begins with fully adjusted weights from the first wave. These weights account for non-response and align with population totals. After each wave's non-response adjustments, population weighting must be reapplied to ensure the panel weights remain compatible with the first wave's totals. However, this repeated adjustment can increase the variability of the weights, which is undesirable.
- An alternative approach applies non-response adjustments to the first wave's base weights without population adjustments. Non-response adjustments are then applied to subsequent waves, with population weighting conducted only at the end. This method reduces the risk of increased weight variability.
- For non-response at wave t , all variables collected up to wave $t-1$ are available for adjustment in subsequent waves. However, it is not practical to include all these variables in the adjustment process, so a selection is necessary. While variables from any previous wave can be used, the focus is typically on those collected in the most recent wave. Ideally, the selected adjustment variables should be related to either response propensity or the variables under study.

14.3.4.2 Adjusting for panel re-entry

- Re-entry occurs when individuals participate in the first wave, drop out for one or more waves, and then re-enter at a later point. In such cases, the original attrition calculations become inaccurate and cannot be used as base weights for the panel. To correct this, attrition weights must be recalculated using re-interview information from the wave in which they return. These returning members are treated as present during the 'drop-out' wave. The recalculated weights, known as shadow weights, then serve as the base for subsequent weight calculations.

14.3.5 Fixed Weighting Structures

- Fixed weighting structures involve keeping weights constant over a sequence of months or years. This method is commonly used in domains where indices are computed. The main advantage of fixed weights is the savings in both time and money, as revising weights can be time-consuming and costly, especially if new surveys are needed. However, potential biases may arise if the base periods become outdated or inconsistent.
- To reduce this bias, several methods can be employed:
 - **Annual Chaining:** This method updates weights yearly, eliminating the need for a fixed base period. The reference period for weights is simply the previous year.
 - **Annual Chaining with Current Weights:** This method updates weights annually as current-year data becomes available, replacing original weights from previous years.
- Seasonality complicates the use of indices in a series. Changing monthly weights to reflect seasonal variations can lead to unreasonable measures year-over-year. A single index cannot precisely measure changes month-to-month and year-to-year, as annual weights ignore seasonal effects.
- A statistical solution is **chain indices**, which link two indices that overlap in a period by rescaling one to match the other within the same timeframe. This approach combines indices into a single time series, accommodating changing weights over time.

14.4 References

01. Battaglia, M et al. (2004). *Tips and Tricks for Raking Survey Data (a.k.a. Sample Balancing)*. Cambridge, Massachusetts: American Association for Public Opinion Research.
02. Cox, B. et al. (2011). *Business Survey Methods*. 2nd ed. New York: John Wiley & Sons, Inc. p490.

03. Groves, R. et al. (2011). *Survey Methodology*. 2nd ed. New Jersey: John Wiley & Sons, Inc.
04. Henry, K et al.(2015). *A design effect measure for calibration weighting in single-stage samples*. Canada: Statistics Canada.
05. Himelein, K (2014). *Weight Calculations for Panel Surveys with Sub-sampling and Split-off Tracking*. United States of America: American Statistical Association.
06. Holt, D et al. (1979). Post Stratification. *Journal of the Royal Statistical Society*. 142 (1), p33-46.
07. ILO, IMF, OECD, UNECE, Eurostat and The World Bank (2004). *Consumer Price Index Manual: Theory and Practice*. Geneva: International Labour Office.
08. Kott, P. (2012). *An Introduction to Calibration Weighting for Establishment Surveys*. Rockville, Maryland.
09. Kott, P (2016). Calibration weighting in survey sampling. *Wiley Periodicals*. 8 (1), p8-39. Marsden, P. et al. (2010). *Handbook of Survey Research*. 2nd ed. United Kingdom: Emerald Group Publishing Limited. p227.
10. Marsden, P. et al. (2010). *Handbook of Survey Research*. 2nd ed. United Kingdom: Emerald Group Publishing Limited. p227.
11. Menard, S (2008). *Handbook of Longitudinal Research - Design, Measurement, and Analysis*. London: Elsevier. p168.
12. Pfeffermann, D et al. (2009). *Sample Surveys: Design, Methods and Applications*. The Netherlands: Elsevier. p342.
13. Potter, F (1990). *Methods for extreme weights in sample surveys*. North Carolina: University of North Carolina.
14. Rose, D. (2000). *Researching Social and Economic Change - the uses of household panel studies*. London and New York: Routledge. p103-104.
15. Rothwell, D (1958). Use of Varying Seasonal Weights in Price Index Construction. *Journal of the American Statistical Association*. 53 (281), p66-77.
16. Schonlau, M et al. (2013). The implementation of cross-sectional weights in household panel surveys. *Statistics Surveys*. 7 (1), p37-57.
17. Statistics Canada (2015). *Survey Methodology*. Ottawa, Ontario: Minister responsible for Statistics Canada.
18. Szymkowiak, M (2016). *Use of Calibration in Official Statistics: Design versus Model-Based Calibration - Pros and Cons*. Madrid: European Conference on Quality in Official Statistics.
19. UN Department of Economic and Social Affairs (2005). *Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid Work*. New York: United Nations. p556.
20. UN Department of Economic and Social Affairs (2008). *Designing Household Survey Samples: Practical Guidelines*. New York: United Nations. p110.
21. Walden, G (2006). *Survey Research Methodology 1990-1999*. United States of America: Library of Congress Catalog.

22. Yansaneh, I (2003). Construction and use of sample weights. *United Nations Secretariat: Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys.*
23. Zhang, L (1998). *Post-Stratification and Calibration - A Synthesis.* Norway: Statistics Norway.



15. Consistency Checks

15.1 Scope and Purpose

Validation checks on statistical data are essential to ensure accuracy and consistency, maximising data quality. Especially in large-scale projects, these checks should begin at the initial stages of data collection.

Consistency checks can take various forms. They range from logical controls, easily implemented in electronic surveys, to data benchmarking and specific statistical checks tailored to particular outputs. After data collection, these checks also encompass other technical aspects of data processing, such as coding quality according to standard classifications.

15.2 Principles

15.2.1 Interim Data Analysis

Many research projects yield findings that directly influence decision-making bodies. Conducting provisional data analysis during collection, known as 'interim' analysis, is crucial to ensure high-quality outputs. This analysis helps verify the consistency of sample data, allowing for early corrective actions when needed.

15.2.2 Core Variables

While all survey questions add value to research results, certain questions are particularly significant. These are known as 'core variables' and are defined by the researchers, often outlined in methodological manuals.

In social surveys of individuals and households, core variables typically cover demographic, geographic, and socio-economic information. Business surveys usually include data on turnover, employment, and the number of business units.

Ensuring the accuracy of information collected on core variables is crucial from the earliest stages of data collection. Benchmarking against data sources with information on the same variables helps ensure consistency across different estimates.

15.2.3 Outlier Values and Trends

During data analysis, identifying and handling outliers is essential. Outliers are data points that deviate significantly from other values in the dataset. Detecting outliers involves defining a plausible value range after examining variable distributions.

In addition to outliers, analysing the distribution of time-recorded values for specific variables can reveal unusual trends. This trend analysis helps identify any anomalies within the data.

15.2.4 'Warning' and Errors

Electronic questionnaires can signal errors to interviewers during data collection. For paper questionnaires, errors can be flagged during data entry.

- **Warnings:** These are errors that interviewers or data entry operators can accept after verifying with respondents or checking the paper questionnaire values.
- **Errors:** These errors are unacceptable, as they indicate clearly illogical responses.

Other errors may arise from the interview software, such as issues with skip patterns or incorrect calculations. While such errors should be minimised through thorough pre-fieldwork testing, early detection is essential if they occur.

15.3 Recommendations

15.3.1 Error Detection

Human error often leads to invalid data. These errors can occur despite the best intentions of users.

Consistency checks during data collection require prior knowledge of which errors are considered 'warnings' and which are 'fatal.' Fatal errors should never appear in the final dataset. Dubious data flagged by warnings should be reviewed carefully. This helps determine if the interviewer or data entry operator conducted necessary checks before accepting a potentially implausible response. Monitoring acceptance rates for such warnings by interviewers or data entry operators is essential.

During data collection, implausible values can be corrected by referencing administrative data sources or other relevant data available within the same context.

15.3.1.1 Outlier detection

The *Tukey* outlier detection method is one of the most widely used and is based on the interquartile range (IQR) of a dataset. As a first step, the first (Q1) and third quartiles (Q3) are calculated, which represent the 25th and 75th percentiles of the data. The IQR is then found by subtracting Q1 from Q3. Outliers are defined as values outside the following bounds:

Lower Bound: $Q1 - 1.5 \times IQR$ and **Upper Bound:** $Q3 + 1.5 \times IQR$

Instead, a value of 3 may be used instead of 1.5, as a larger value results in stricter criteria for identifying outliers. Any values outside this range are considered outliers. This method is effective for detecting outliers in skewed or non-normal distributions and doesn't rely on assumptions about the data's underlying distribution.

15.3.2 Checks for Consistency and Benchmarking of Data

For surveys conducted over several weeks, early consistency checks are essential. These checks detect flaws or misclassification errors initially, enabling timely corrective actions.

Demographic, geographic, and socio-economic data should be cross-checked with reliable estimates from other sources. These estimates provide the foundation for accurate statistics on the topic. Core socio-economic figures should be analysed alongside key demographic and geographic details, such as gender, age, and district of residence.

Business demographic variables—such as the number of units, turnover, and employment—and specific macroeconomic indicators must be benchmarked against the NSO's Statistical Business Register and estimates from the SBS Survey. SBS statistics cover industry, trade, and services, offering insights into the structure and performance of businesses. Information from business units should be validated against results from these sources.

When a population and housing census is conducted every ten years, it is crucial to ensure that census figures align with estimates from related domains for the same reference year. This alignment enables future extrapolations based on census data. Census data, representing a national count of people residing in Malta, helps eliminate sample-related biases by providing a comprehensive baseline.

15.3.2.1 Labour Status

Information on the number of employed, unemployed, and economically inactive individuals should be benchmarked against estimates from the LFS. The LFS provides regular and harmonised data on the labour market.

15.3.2.2 Employment Status

The distribution of employed individuals by employment status—distinguishing between employees, employers, and own-account workers—should be cross-checked with LFS estimates.

15.3.2.3 Highest Level of Education Completed

Statistics on educational attainment need verification against LFS estimates. The classification of educational levels should follow ISCED guidelines, as directed by the NSO Unit responsible for education statistics.

15.3.2.4 Household Income

Household income data should be benchmarked with the Statistics on Income and Living Conditions (EU-SILC) results. This survey provides comparable data on income, poverty, and social exclusion.

15.3.2.5 Citizenship

Data on foreigners without Maltese citizenship should be cross-verified with population statistics estimates.

15.3.2.6 Household Composition and Size

Data on the distribution of private households by size and estimates of average household size should be benchmarked against population statistics.

15.3.2.7 District of Residence

Household distributions by geographical district of residence should be cross-checked with population statistics to ensure accuracy.

15.4 References

01. Bureau of the Census (1992). *1992 Annual Research Conference: Proceedings*. Virginia, USA: Census Bureau.
02. United Nations Economic Commission for Europe (2006). *Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing*. Geneva: UNECE.

15.5 Other Sources

The Art of Data Validation. Written by S. Richard. Available at: <https://www.petesqbsite.com/sections/express/issue10/datavalidation>.



16. Time Series and Seasonal Adjustment

16.1 Scope and Purpose

In various statistical fields, data is often recorded in a time series, which allows the assessment of economic trends over time. Time series data, collected at equally spaced intervals, captures seasonal variations. This format is useful for analysing long-term economic behaviour and tracking labour market indicators.

A trend reflects a long-term increase or decrease in the data. In contrast, seasonal variations are recurring factors that repeat over time. Seasonal fluctuations and calendar effects can obscure patterns in a time series, making it difficult to understand underlying trends. Therefore, filtering out regular seasonal variations and calendar effects is essential. By removing these, data comparisons across different months or quarters become possible. Seasonally adjusted data also enable comparisons between series with different seasonal patterns, facilitating international economic comparisons.

In STBS, most indicators exhibit seasonal patterns due to business cycle influences. Correcting time series data for seasonal fluctuations and calendar effects helps clarify the underlying economic developments. The NSO submits STBS indicators in a seasonally adjusted format to meet European Statistical System requirements. Similarly, the labour cost index (LCI) components—such as wages, social security contributions, and total labour costs—are transmitted in adjusted forms (working-day and seasonally adjusted). National Accounts data is also submitted in a seasonally adjusted format.

16.2 Principles

16.2.1 Time Series Components

Time series analysis reveals that data observations evolve through various patterns. These patterns include trends, cycles, seasonal variations, and irregular fluctuations. Collectively, these are known as time series components.

- **Trend:** This represents long-term variations. Trends may show steady growth or consistent downward movement over an extended period.
- **Cycle:** Cycles occur over shorter intervals compared to trends and involve periodic oscillations. They reflect alternating periods of growth and contraction.
- **Seasonal Fluctuations:** These are systematic changes that recur regularly, often due to climatic or institutional factors. Seasonal movements typically repeat with similar intensity each year during the same season and can be anticipated based on historical patterns.
- **Irregular Components:** These are random, unexpected movements, such as those caused by floods, strikes, or unseasonal weather. Irregular components do not follow a predictable pattern.

16.2.2 Decomposition

Analysing a time series involves identifying its components and understanding their nature. A time series can be decomposed into trend, cyclical, seasonal, and irregular components. This process illustrates how these components combine to form the original series.

Decomposition can follow either an additive or a multiplicative model. Multiplicative models are commonly used, as the seasonal component varies proportionally with the series level. Additive models are suitable when the seasonal component remains independent of the series level.

16.2.3 Seasonal Adjustment

Seasonal adjustment removes seasonal fluctuations from a time series, resulting in seasonally adjusted data. This process is essential for observing changes over time and gaining a clearer view of key indicators. An apparent sharp increase in unadjusted data may prove insignificant once seasonally adjusted.

There are two main methods for seasonal adjustment: *TRAMO/SEATS* and the *X-11/X-12* or *X-13* approach. Additionally, mixed seasonal models with moving average parameters may be applied. The *TRAMO/SEATS* method is commonly used by various domain units within the NSO.

16.2.4 Pre-treatment

The main goal of pre-treatment is to ensure reliable estimation of the seasonal and calendar components in a time series. This involves detecting and correcting for outliers and calendar effects.

Outliers can be visually identified through basic graphical analysis, which reveals abnormal values or spikes. Alternatively, automatic outlier correction, using specific software such as *JDemetra+*, may be used. *JDemetra+* identifies various outliers, including additive outliers (AO), transitory changes (TC), and level shifts (LS). Once identified, outliers are corrected in the unadjusted data before pre-treatment.

In **STBS**, automatic pre-treatment is applied to all time series submitted to Eurostat or published locally. The software automatically detects outliers, and no manual analysis of outliers is typically performed. For the **LCI**, pre-treatment includes an initial graphical analysis of both unadjusted data and the original time series graph.

In **National Accounts**, *JDemetra+* also detects outliers automatically, but some manual intervention occurs. Analysts inspect the original series graph to validate outliers identified by the software. Each confirmed outlier is supported with an economic interpretation.

Calendar effects arise from variations in the number of working or trading days in a month or quarter or from the timing of public holidays. Common calendar effects include differences in working days, working day composition, leap year effects, and moving holidays. Effects related to bridging days, school holidays, and weather are generally not considered, as they tend to neutralise over time. Calendar adjustment is necessary when statistical evidence supports calendar effects in a time series. To correct for calendar effects, the national Maltese calendar is applied. The National Accounts Unit uses three types of calendars, depending on the time series analysed:

- **General Calendar (Excluding Bank Holidays):** This calendar includes Maltese public holidays but excludes bank holidays. It is used for all series where NACE Section K (Financial and Insurance Activities) is not included.
- **Calendar for NACE Section K (Including Bank Holiday):** This calendar includes all Maltese holidays plus an additional bank holiday on 26th December. It applies to the NACE Section K series, as entities in this sector observe an extra holiday. This holiday is weighted at 1.

- **Adjusted Calendar for NACE Section K (Weighted Bank Holiday):** This calendar includes the additional holiday for NACE Section K but with a weight lower than 1. This adjustment is based on observed estimates of the full-time equivalent workforce in NACE Section K and is used for other series influenced by this sector.

16.2.5 Graphical Analysis

Seasonality can often be identified early in the analysis by viewing S-I ratio graphs. These charts are useful for examining seasonal patterns, helping to detect unstable or moving seasonal factors. They display the final and mean seasonal factors for each time interval.

Graphical analysis also reveals the structure of the series. Sometimes, a visual inspection can indicate whether a series is seasonal. However, this may not be clear in problematic series or those with complex shapes.

The graph's shape can also suggest whether the series should be split. It may reveal seasonality breaks or indicate errors in raw data or method changes between years. The raw data is corrected at the source if an input error is detected.

16.2.6 Revisions

Revisions to seasonally adjusted time series occur for two reasons: changes in the unadjusted data or improved identification of seasonal patterns as new observations become available.

In STBS and National Accounts, the partial-concurrent adjustment method is applied. This approach involves re-identifying the model, filters, outliers, and calendar regressors once a year. The respective parameters and factors are re-estimated whenever new or revised data becomes available. Any revisions in unadjusted or seasonally adjusted data are incorporated, and the entire time series is revised each time new information is received. The time series undergoes revisions for both routine updates and benchmark revisions. Additionally, the National Accounts Unit applies the 'partial-concurrent adjustment fixed models' procedure with each quarterly news release. This method keeps the model and identifies outliers fixed for the subsequent years until the next annual seasonal adjustment is performed.

LCI data are revised quarterly to incorporate changes from primary data sources. The last three years remain open to revisions, as new survey information may update data on wages, salaries, and social contributions.

For National Accounts, revisions to unadjusted and seasonally adjusted data may occur between any two quarterly releases. Additionally, annual revisions are conducted when all series undergo a comprehensive assessment.

16.3 Recommendations

16.3.1 Justifying Seasonal Adjustment

- Apply seasonal adjustment only when seasonal and/or calendar effects are identifiable, explainable, and estimable.
- If no seasonal or calendar effects are present in the data, the unadjusted series will equal the calendar/seasonally adjusted series.

16.3.2 Outlier Detection

- For marginally significant outliers, assess whether to retain them in the model. Do not include any outliers without a valid reason.
- Automatic outlier correction results should be verified by reviewing graphs of the series and considering any relevant information (economic, social, etc.) that could explain the detected outlier. Outliers with an economic explanation, such as strikes, changes in government policy, or territory changes affecting countries or economic areas, should be retained.

- Include outliers with a clear interpretation as regressors in the model, even if their effects are slightly below the usual significance threshold.
- Verify outliers detected automatically by software through graph inspection and data verification.
- Outliers at the end of the series (in the last year) should be carefully evaluated. If an economic interpretation confirms them, classify them as AO.
- For outliers detected at the first data point, whether automatically or by inspection, include them only if they have an economic interpretation and are statistically significant. Classify them as AO or LS rather than TC.

16.3.3 Validating Seasonal Adjustment

- The observed time series should ideally span more than four years.
- Before publishing seasonally adjusted results, analyse the figures to confirm their validity.
- Review the results of seasonality tests. JDemetra+ automatically outputs six seasonality tests: auto-correlations at seasonal lags, Friedman (non-parametric), Kruskal-Wallis (non-parametric), spectral peaks, periodogram, and seasonal dummies. The auto-correlation test at seasonal lags is preferred for LCI components. In National Accounts, the results of the Friedman and Kruskal-Wallis tests indicate whether the time series is seasonal.
- If seasonality tests yield inconclusive results, confirm seasonality by analysing decomposition. This involves presenting the relative contribution of the components to the stationary portion of the variance in the original series after removing the long-term trend. If the seasonal component accounts for 25% or more, the series is confirmed as seasonal and should be adjusted. If not, it is recommended that the series not undergo seasonal adjustment.
- If the test statistics indicate issues with normality or diagnostics for model residuals, adjust the regressors to ensure the model constraints are respected, especially regarding working day or leap year effects.
- Once tests confirm the validity of the seasonal adjustment, compare the current seasonally adjusted series with the previous one to assess the magnitude and length of revisions. If large, consistent revisions occur throughout the series, flag the series and re-estimate for confirmation.
- Graphical analysis provides crucial information about the validity of seasonal adjustment. It is essential to interpret identified outliers; if this is not possible, the statistician must ensure no issues with the raw data.
- The decomposition model should be examined. A model fits the data well when the seasonal component is significantly stronger than the irregular component.
- Reapply seasonal adjustment until reliable results are achieved.

16.3.4 Revisions

- Revisions to seasonally adjusted data must be published according to official revision guidelines, including the revisions policy and calendar.
- Revised seasonally adjusted data should not be released more frequently than unadjusted data.
- Inform users about past revisions' size, direction, and volatility for important seasonally adjusted macroeconomic variables.

16.3.5 Documentation and Archiving

- Systematically save and store all data (unadjusted, calendar, and seasonally adjusted) for each time series and period.
- Additionally, store revision analyses and their metadata for each time series and period for future reference or analysis.

16.4 References

01. Bianconcini, S. *et al.* (2016). *Seasonal Adjustment Methods and Real-time Trend-cycle Estimation*. Bologna (University of): Springer.
02. Boxall, M. *et al.* (2015). *ESS Guidelines on Seasonal Adjustment*. Luxembourg: Eurostat.

03. Donovan, R.J. and Norwood, J.L. (1982). *A Guide to Seasonal Adjustment of Labour Force Data*, Bulletin No. 2114. Washington D. C.: United States Bureau of Labour Statistics.
04. Yaffee, R. *et al.* (2000). *Introduction to Time Series Analysis and Forecasting, with Applications of SAS and SPSS*. New York: Academic Press.



17. Assignment of Disposition Codes and Response Rate Calculation

17.1 Scope and Purpose

The success of any survey heavily depends on achieving an adequate response rate. A high response rate ensures that the total population is better represented, allowing for more robust calculations and reducing bias in estimates.

Non-response can have two significant effects on the data. First, if non-respondents differ from respondents, estimates may become biased. Second, the overall variance of estimates increases due to the reduced sample size. These factors highlight the importance of accurately estimating response rates.

Assigning final disposition codes to each sampling unit is crucial for calculating response rates, which helps clarify non-response errors. Unit non-response occurs when a survey fails to collect data from selected units. Non-responding sampling units are eligible but provide no information, or the information collected is unusable.

17.2 Principles

17.2.1 Main Groups of Survey Cases

Each sampling unit in the sampling frame can be classified and coded as eligible, ineligible, or with unknown eligibility. Eligible units include both respondents and non-respondents.

17.2.2 Response Rate Calculation

Response rates can be calculated after categorising the sample elements. Two types of response rates can be determined: unweighted and weighted.

- **Unweighted Response Rate:** This measures the proportion of the sample that provided usable information for analysis, specifically the eligible units that cooperated in the survey.
- **Weighted Response Rate:** This indicates the proportion of population elements that would have been respondents if asked to participate.

The unweighted response rate R_{uw} is calculated as follows:

$$R_{uw} = \frac{n_R}{n_R + n_N + n_U}$$

Whereas the weighted response rate R_w is calculated as follows:

$$R_w = \frac{\sum_R d_j}{\sum_R d_j + \sum_N d_j + \sum_U d_j}$$

Where:

n_R is the number of responding units

n_N is the number of eligible non-responding units

n_U is the number of units with unknown eligibility⁷

d_j is the design weight of the unit in the sample (inverse of inclusion probability)

$\sum_R d_j$ is the sum of weights attributed to responding units

$\sum_N d_j$ is the sum of weights attributed to eligible non-responding units

$\sum_U d_j$ is the sum of weights attributed to units with unknown eligibility

Consequently, the unweighted non-response rate is $1 - R_{uw}$, and the weighted non-response rate is $1 - R_w$.

17.3 Recommendations

Assign survey disposition codes as outlined below, considering eligibility, item non-response, and the mode of data collection.

17.3.1 Response

17.3.1.1 Eligible Unit – Good Response

- This code applies to units that provided all required data and those that partially completed the questionnaire.
- Partially filled questionnaires are classified as good responses if they include all essential basic information about the main indicators. The list of main indicators varies by domain and type of survey. The questionnaire remains valid if information on these indicators can be obtained from administrative data sources.

17.3.2 Non-response

17.3.2.1 Eligible Unit – Refusal

- This code applies to sampled households, individuals, or enterprises that refused to participate in the survey for various reasons, whether outright or circumstantial.
- Common reasons for refusal include lack of time, perceptions of the survey being too extensive, lack of interest, and feelings of privacy violation.
- This category also includes households, individuals, or enterprises that initially agreed to provide information later (with an appointment set) but subsequently refused or could not be contacted.
- Partially filled questionnaires that lack all necessary basic information should be classified as refusals, referred to as 'break-offs'.

17.3.2.2 Eligible Unit – Non-contact

- This code applies to sampling units where eligibility is confirmed through external sources, but contact cannot be established.
- In face-to-face and telephone surveys, non-contact occurs when sampled households or individuals are temporarily unavailable during fieldwork (for example, if they are not home or abroad). Households without a doorbell at the main entrance of an apartment block should also be classified as non-contact.

⁷ A cautious approach is adopted, where all units with unknown eligibility are included in the calculation. Unless there is compelling evidence to suggest that only a specific proportion of these units are actually eligible, it is advisable to assume that all are eligible.

- Unsuccessful telephone attempts, such as no answer, busy signals, or answering machines, are also considered non-contact cases.
- In self-completion surveys, non-contact units never returned the paper or web survey.

17.3.2.3 Eligible Unit – Other Reasons

- Classify households or individuals as ‘other eligible’ when communication is impossible due to linguistic problems.
- Similarly, code respondents who cannot communicate because of physical incapacity, old age, physical or mental illness, other health issues, or family problems as ‘other eligible’.
- Individuals who died during or after the reference period should also be classified as ‘other eligible’.

17.3.2.4 Ineligible Units

- The ineligibility of units is determined through direct contact or external information, such as data from neighbours or linkable databases.
- In face-to-face surveys, addresses are considered ineligible if they do not exist or are non-residential (e.g., a car garage). Due to incorrect details, this classification also applies to vacant dwellings, summer residences, properties under refurbishment, and unreachable addresses.
- Individuals who died before or on the first day of the reference period are regarded as ineligible.
- Households with all members residing in institutions and individuals living in institutional settings should be recorded as ineligible.
- In business surveys, ineligibility applies to units that have ceased operations, those temporarily inactive, out-of-scope units (e.g., due to misclassification), and units that no longer operate independently following a merger or takeover.
- In postal surveys, ineligible units include undelivered surveys marked by postal workers as ‘moved away.’ This also applies to undelivered emails in email surveys.
- In telephone surveys, ineligible units refer to unreachable individuals due to incorrect phone numbers, disconnected lines, non-working numbers, faxes, or answering machines.
- Telephone numbers sampled but not attempted before the end of the field period (due to quota limits) should also be coded as ineligible.

17.3.2.5 Units with Unknown Eligibility

- Units with unknown eligibility are those for which it is impossible to determine the eligibility of a sample unit. These cases typically arise when contact cannot be established or eligibility information cannot be retrieved from external sources. Whether an eligible residential household or individual exists at the sampled telephone number or address remains unclear.
- Addresses that exist but cannot be reached due to their location on the outskirts or in inaccessible areas should be coded as unknown eligibility.
- Similarly, if external sources confirm that foreigners inhabit a sampled address without confirmed usual residence in Malta, these should also be recorded as units with unknown eligibility.
- In self-completion surveys, this code applies to undelivered postal surveys for which the postal worker indicated a reason other than ‘moved away’.

17.3.3 Response Rate Calculation

- Response rates must be calculated for every survey using the formulas outlined above.
- In surveys of establishments, a small number of large establishments may represent a significant portion of the total population. In these cases, the weighted response rate is typically used to indicate the quality of the estimates.

17.4 References

01. McCarty, C. (2003). Differences in response rates using most recent versus final dispositions in telephone surveys. University of Florida.



18. Survey non-Response

18.1 Scope and Purpose

No matter how well a project is planned, challenges may arise, and data collection is no exception. Despite all efforts, some level of non-response is inevitable in any survey. Producing high-quality statistics requires high response rates and low sampling variances, achievable through careful design of statistical processes and consideration of the target population.

Maximising response rates in a survey is crucial for accurate and representative data collection. Non-response affects data quality, so measures must be taken to reduce and correct it. Non-response correction can be achieved using statistical imputation techniques and weighting procedures.

18.2 Principles

18.2.1 Types of Non-response

Non-response in surveys can be classified into unit non-response and item non-response.

- **Unit Non-Response:** Also known as first-level non-response, this occurs when data is unavailable for an entire statistical unit.
- **Item Non-Response:** Also referred to as second-level non-response, this happens when a unit participates in a survey, but data for certain items is missing. Item non-response can manifest in different patterns:
 - **Partial Non-Response:** Data is missing from a specific point in the questionnaire.
 - **Random Missing Data:** Some items are missing without a clear pattern.
 - **Missing by Design:** Some items are intentionally left out due to logic skip routines, where not all items apply to every respondent category.

18.2.2 Reasons for Non-response

Unit Non-Response occurs when a statistical unit cannot be contacted or refuses to participate. Non-contact typically happens when an interviewer cannot locate the selected individual in face-to-face surveys. Other examples include undelivered postal questionnaires and no response in telephone surveys.

Partial Non-Response arises when an interview is interrupted. This is common in telephone surveys, where the respondent may refuse to continue, or the call may be disconnected. In panel surveys, respondents may drop out during specific waves, a phenomenon known as panel attrition.

Item non-response can occur for several reasons. A question may be overlooked or deliberately left unanswered if considered

sensitive. In some cases, respondents may genuinely not know the answer. Responses such as 'do not know' should be handled carefully, depending on the nature of the research. In certain surveys, 'do not know' can be a valid response.

Item non-response may also occur even when an answer is provided. For example, a response might fall outside the acceptable range or be unreadable in self-administered surveys. Data processing errors, such as incorrect data entry, can also result in unusable information.

18.3 Recommendations

Several measures can be employed during the data collection to ensure better response rates.

18.3.1 Advance Letters

- One effective method to reduce unit non-response is sending advance letters to selected individuals or households before collecting survey data. This approach is particularly useful for face-to-face interviews.
- Advance letters should provide key details, including the survey's purpose, the methodology behind the selection process, how confidential information will be handled, and when contact is expected.
- The timing of these letters is crucial. Sending them too early may lead respondents to forget about the survey. Sending them too close to the data collection period may result in late delivery. Ideally, advance letters should be sent two to three days before interviewers schedule appointments with respondents.

18.3.2 Mode of Data Collection

- The mode of data collection should suit the specific survey. For instance, while self-administered electronic questionnaires may work well for businesses, they may not be ideal for individuals without easy internet access. Self-administered questionnaires are preferred when it is beneficial for respondents to answer at their convenience.

18.3.3 Proper Management of Interviewers and Interviewing Period

- Sufficient time should be allocated for data collection, allowing interviewers to revisit households or schedule appointments for visits or calls. Although logistics and time constraints may limit this period, extending the survey duration can help reduce unit non-response. A longer survey period generally leads to higher response rates.
- Interviewer workloads must be managed carefully. Excessive assignments may cause interviewers to rush their tasks or neglect proper follow-up. Ensuring they have adequate time for each case improves data quality.
- Interviewers should leave 'non-contact forms' at households where respondents are unavailable. These forms provide contact details for scheduling an alternative appointment. Interviewers should attempt to contact the household up to three times on different dates and times.
- A clear deadline for survey completion must be communicated to potential respondents, especially in postal surveys, which are self-administered. Questionnaires sent by post should include an informed consent form that provides respondents with key information, such as the purpose of the survey, how their participation supports the survey's goals, assurance of anonymity, and the deadline for returning the completed questionnaire.
- The time allowed for survey completion depends on the project timeline and specific requirements like transmission or results deadlines. Survey duration should be carefully considered. Respondents should be given at least two weeks to complete the questionnaire. The data collection period for business surveys can extend up to four weeks.
- Reminders effectively encourage responses for self-administered questionnaires, depending on the time available for the survey. The timing of reminders should account for the survey duration and the delivery method. For instance, a postal reminder takes longer than a phone reminder to reach respondents.
- Reminders should follow the same format as the original survey invitation, clearly highlighting the survey title and scope. Both the original deadline and the extended due date should be emphasised. For postal surveys, another copy of the questionnaire and a pre-paid, self-addressed envelope should be provided upon request.

- A clear reminder schedule is essential for managing non-response. For a four-week survey, the first reminder should be sent two weeks after the initial invitation. A second reminder should follow a week later. Non-respondents should be contacted by phone two to three days after the survey deadline. The number of reminders should be determined based on the extent of unit non-response. For surveys distributed via email, reminders should be sent the day after the established deadline.

18.3.4 Skilled Interviewers and Proper Training

- The skills of interviewers play a crucial role in reducing item non-response. If an interviewer fails to state or explain a question clearly, the respondent may refuse to answer or provide an incorrect response.
- Good interviewing is not just about hiring people with natural abilities. Interviewers need regular training, and briefing sessions should be specific to each survey. These sessions should cover the survey's purpose, scope, and potential challenges.

18.3.5 Assuring Confidentiality

- Respondents may hesitate to participate or provide honest answers if they have concerns about the confidentiality of their information. Therefore, it is essential to implement strong measures that reassure them their data will be properly protected. These measures should include the interviewers' professionalism and the surveying organisation's integrity.

18.3.6 Incentives

- Offering incentives is another way to increase response rates for fully completed questionnaires. Although this approach can be costly, it may be necessary, especially for long and tedious surveys.
- Care must be taken to ensure that the incentive does not disproportionately appeal to a specific subgroup, which could skew the survey results.
- One possible incentive is to provide participants with a copy of the aggregate survey results. This allows respondents to see the value of their input. Additionally, anonymised microdata from the survey can be made available to recognised researchers. However, it must be clearly stated that such data will not be provided for general knowledge purposes.

18.4 References

01. Berkenbosh, B. (2011). 'Data Collection Strategy: Response-enhancing Measures for Business Surveys'. *Statistical Methods 201109* (paper). The Hague: Statistics Netherlands.
02. De Leeuw, E. D., Hox, J. and Huisman, M. (2003). 'Prevention and Treatment of Item Non-response'. *Journal of Official Statistics* 19 (2): 153-176.
03. Luiten, A. (2011). 'Data Collection Strategy: Response-enhancing Measures for Social Statistics'. *Statistical Methods 201121* (paper). The Hague: Statistics Netherlands.



19. Presentation of Statistical Data

19.1 Scope and Purpose

The presentation of statistical data involves publishing information collected from various sources. This allows detailed information to be presented concisely. Results can be illustrated through news releases, publications, presentations, and online media such as articles, interactive charts, and maps.

Statistical data can be presented in three main formats, namely text, tables and charts (or diagrams, graphs and maps). The choice of format depends on the volume, type, and complexity of the data, as well as the target audience. Text, tables, and charts should be easy to interpret and well-organised. Tables and charts must clearly convey what the data shows without repeating the information in a text.

Text is typically used to explain findings, outline trends, and provide context. Tables are ideal for presenting individual characteristics and displaying both quantitative and qualitative information. Graphs, on the other hand, are better suited for illustrating trends, making comparisons, and showing relationships. They visually reveal data patterns, such as changes over time and frequency distributions.

Well-presented tables and graphs can summarise information that would otherwise be lengthy. Poorly-presented tables and graphs, however, can confuse and mislead

19.2 Principles

19.2.1 Tables

Tables are a set of figures systematically displayed row-wise and column-wise. These are self-explanatory and should be able to be read without reference to text.

19.2.2 Graphs and Charts

- **Line Graphs** present time series (scale) data, with the x-axis representing time. A line graph connects data points to show continuous changes over time. They are particularly useful for identifying patterns and trends, such as seasonal effects, significant changes, and turning points.
- **Bar Charts** are ideal for displaying categorical data. Rectangular bars have a height (or length) proportional to the values they represent and can be plotted vertically or horizontally. In vertical bar charts, the x-axis represents the different categories without a scale, while the y-axis has a scale indicating units of measurement. Vertical bar charts are preferred for comparing fewer than eight groups, while horizontal bar charts are better for comparing eight or more groups.
- **Histograms** are a type of bar chart where the area of each rectangle is proportional to the frequency of a variable, and the width corresponds to the class interval. Unlike standard bar charts, histograms represent continuous data rather

than discrete categories. Since scale data may have many possible values, it is often grouped to reduce the number of data points.

- **Pie Charts** are circular graphs where each slice's arc length and area represent the quantity's proportion. They are commonly used to display percentage distributions of a variable.
- **Radar Charts** (web or spider charts) display multivariate data in a two-dimensional chart with three or more quantitative variables. These variables are represented on axes that start from the same central point. The relative position and angle of the axes are typically not informative. A node on the axis shows each variable's value, and lines connect these nodes to form a polygon.
- **Maps** are diagrammatic representations of land or sea areas. **Choropleth maps** are thematic maps where areas are shaded or patterned in proportion to the measurement of the displayed statistical variable.

19.3 Recommendations

19.3.1 Textual

- Use the same font face, colours, and sizes consistently throughout the release or publication.
- Avoid including more than three or four numbers in a single sentence. Instead, use tables or charts to present larger sets of numerical data.
- Write whole numbers less than ten as words, and use digits for numbers 10 and above.
- Reference tables in text using brackets following the word 'see' (for example, 'see Table 1').
- Use commas to separate figures in large numbers.
- Avoid placing two numbers next to each other in a sentence unless one is in parentheses.
- Write fractions as words.
- Use figures for decimal numbers, rates, percentages, dates, and ages.
- The terms 'ratio,' 'proportion,' 'percentage,' and 'rate' are often incorrectly interchanged. A **ratio** expresses one number concerning another by dividing one by the other. A **proportion** is a specific type of ratio where the denominator is the total. A **percentage** is a proportion based on 100 units. **Rates** are another ratio form, representing the probability of a certain event.
- Use the term '**percentage points**' when comparing two percentages. For example, an increase from 10 per cent to 15 per cent is an increase of five percentage points. Referring to this as a 5 per cent increase is incorrect, as it implies a 5 per cent rise on the original 10 per cent, resulting in 10.5 per cent rather than 15 per cent.
- Ensure any definitions and units used in data presentation are clearly visible.
- Describe an increase as 'an increase over' rather than 'an increase from'. For decreases, use 'decrease from'.
- Use 'preceding' instead of 'previous' when referring to something that occurred immediately before.
- When describing numbers, use 'larger/largest' or 'smaller/smallest.' For rates, use 'higher/highest' or 'lower/lowest.'
- Use 'per cent' in written reports instead of the '%' symbol.
- Commentary on tables or graphs should provide enough detail to be informative but remain simple for the reader. Avoid cluttering the text with excessive figures; use an additional table if needed for clarity.

19.3.2 Notes, Footnotes, and Sources

Notes provide general information that applies to all categories in a table. These should be placed directly under the table.

- Footnotes offer specific details about particular categories, such as a break in series for certain countries. Footnotes should follow the notes and be numbered according to the text. Numbers are preferred over the asterisk (*) symbol.
- Sources of the data should be clearly stated, following the notes and footnotes.
- The following codes can be used to describe the status of figures in tables:
 - o (c) confidential
 - o (:) not available
 - o (p) provisional
 - o (u) unreliable

19.3.3 Table and Chart Titles

- Titles should be short, accurate, and easily understandable, not exceeding 150 characters.
- The title should begin with key content, such as 'population,' 'enterprises,' or 'expenditure.'
- This should be followed by the necessary breakdowns, such as 'by sex,' 'age,' 'marital status,' 'educational level,' 'economic activity,' or 'size class'. Breakdowns should be introduced with the word 'by'.
- Column-wise breakdowns should appear before row-wise breakdowns.
- Common acronyms, like 'GDP', do not need to be spelt out. Less well-known acronyms should be written in full, with abbreviations in brackets.
- The specific version of any standard classification should be stated, for example, 'NACE Rev. 2 activity.'
- International symbols for metric units, such as 'kg' and 'km²', should be included in the titles where relevant.
- Denominations of figures, such as 'thousands,' 'millions' or 'billions', should be clearly stated in the titles. The same applies to monetary units, for example, 'thousand EUR'.
- Use the '%' symbol for percentages.
- Only commas, hyphens, colons, and round brackets are permitted in titles, and they should be used appropriately to improve clarity.
- Use the word 'and' instead of the ampersand '&', except when the ampersand is part of a universally recognised term, such as 'R&D.'

19.3.4 Tables – Good Practice

- Publications should include a referenced list of tables.
- Figures in tables must have consistent decimal places and use commas for separation.
- The number of digits and decimal places should suit the table's purpose. Use the minimum necessary digits to present all information, ensuring rounding is mathematically correct.
- For tables with percentages, round to one decimal place. More than one decimal place may be required for indicators showing monthly changes.
- Totals should be calculated using unrounded figures and then rounded afterwards.
- Large numbers should be rounded to thousands, millions, or billions, as appropriate.
- Add disclaimers to avoid inconsistencies.
- Tables must be additive both horizontally and vertically. If totals do not tally due to rounding errors a note should be included.
- Items in table cells should be as homogeneous as possible, ideally representing values of a single variable measured consistently.
- Use percentages that add up to 100 across rows, columns, or the entire table to facilitate comparisons in rows or columns with large counts. Averages can also be used.
- Ensure the order of rows and columns follows a logical sequence. In a series of tables with the same rows or columns, maintain consistency across all tables.
- Avoid large gaps in columns, as they make tables harder to read.
- Detailed tables should be summarised in the main text, with the full table annexed.

19.3.5 Graphs and Charts – Good Practice

- Graphs should visualise key findings with simplicity, making them more effective.
- Build graphs from original data using unrounded figures.
- If displaying more than one variable, use legends or keys to differentiate between them.
- Label both axes and indicate units of value and/or periods.
- Remove the y-axis line and include horizontal gridlines in light grey.
- Draw horizontal gridlines at regular intervals based on the major divisions of the y-axis scale, ensuring they are labelled.
- Display the denomination on the y-axis (e.g., 'EUR', '000') above the highest figure on the y-axis.
- For a monthly time series, show months by their initial letter, with the relevant year displayed underneath.

- For a quarterly time series, label the quarters as Q1, Q2, etc., with the year underneath.
- Use gridlines to help users distinguish between months, quarters, and years.
- When shading is used, include a key or legend on the graph.
- Place chart legends below the chart, with items displayed side by side. For pie charts, position the legend on the right-hand side. If the legend text is too long or there are too many series, display them below the chart in vertical order.
- Avoid using three-dimensional graphs, as they can be difficult for many users to understand.

19.3.5.1 Line Graphs

- Line graphs are ideal for presenting time series (scale variables), effectively showing variations and trends. Avoid using smooth lines if the data represents measured values.
- Limit line graphs to a maximum of five lines.
- Distinguish lines using different colours, plotting symbols, or line styles.
- Use a consistent scheme of colours, line styles, and plotting symbols throughout the same publication.
- Comparisons within line graphs are only valid if all lines use the same scale.
- The vertical axis does not need to start at the origin.

19.3.5.2 Horizontal Bar Graphs

- Horizontal bar graphs are ideal for displaying eight or more categories.
- They work well when category labels are too long to fit neatly under the x-axis.

19.3.5.3 Histograms

- Bars in histograms should be arranged in ascending or descending order based on the range of values.
- Each bar represents a range of data, and its width should be proportional to the width of each category. Bar height must correspond to the frequency or percentage of the respective category.

19.3.5.4 Pie Charts

- Pie charts display one variable at a time for categorical data. If comparing two or more variables or showing changes over time, create multiple pie charts and present them side by side.
- A pie chart should not have more than six segments.
- A legend is unnecessary if category names, values, or percentages are written directly in the segments. If needed, place the legend on the right side of the chart.
- Segments should be arranged from the largest to the smallest, starting at 12 o'clock.
- Caution should be exercised when using a pie chart to represent data if any categories have significantly lower values compared to the others, as they may be difficult to distinguish.

19.3.5.5 Radar Charts

- In radar charts, all axes must share the same scale and be equidistant.
- Grid lines connecting the axes can provide a helpful guide.
- Limit the chart to no more than eight variables to maintain clarity.

19.3.5.6 Maps

- Maps should have titles similar to other chart types and must include a legend and scale.
- Legends should be clearly labelled and shaded to represent the data range.
- If a table accompanies a map, there is no need to repeat the data source on the map. If the source is not indicated elsewhere, place it at the bottom of the map.

19.4 References

01. Duquia R.P. *et al.* (2014). 'Presenting Data in Tables and Charts'. *An Bras Dermatol* 89 (2):280–285
02. Miah, A.Q. (2016). *Applied Statistics for Social and Management Sciences*. Singapore:Springer.
03. United Nations Economic Commission for Europe (2009). *Making Data Meaningful Part 2: A Guide to Presenting Statistics*. Geneva: UNECE.



20. Revision of Statistical Data

20.1 Scope and Purpose

Published data is often revised due to the availability of additional information, the receipt of amended data, or changes resulting from more detailed data editing. Revisions may also occur as a result of changes in methodology. Revisions are generally defined as changes to statistics already released to the public. Revisions are unavoidable, particularly when statistics on economic developments are produced shortly after the reference period.

The main reason for revisions to statistical figures is late reporting by data providers. Missing values are usually estimated and revised to ensure timely production when the same statistics are re-released. New versions of standardised classifications may also necessitate revisions, as redefined categories result in differences between new statistics and those based on older classifications. Revisions can make comparing statistics difficult, so entire time series often need to be revised. Other sources of revisions include errors in the compilation process or corrections made by respondents to fix inaccurate reporting.

From the users' perspective, revisions can be beneficial as they improve the quality of available information. However, changes to economic data, for example, may lead to a different assessment of the economy and cast doubt on past empirical results. From the producers' perspective, while additional information leads to more accurate statistics, frequent revisions can undermine the credibility of the data.

20.2 Principles

20.2.1 Main Types of Revisions

As specified in CDU-POL-002: Revisions of Official Statistics Policy, revisions can be classified into three main types: routine, benchmark (methodological), and non-scheduled.

- **Routine revisions** are changes to data that occur at regular intervals, usually on pre-scheduled dates. For example, in monthly news releases, the revised results of the previous month are typically published alongside the new release. Routine revisions mainly involve updating the data to include more complete and current information. Routine revisions typically apply to statistical domains that combine data from various sources. These revisions involve regular updates to improve the completeness of the data.
- **Routine ad-hoc revisions** primarily stem from the integration of new data sources, which replace previous estimates with more accurate figures, and/or potential methodological updates that may lead to more substantial changes than those associated with standard routine revisions. The Office ensures advance communication of routine ad-hoc revisions to its users.
- **Benchmark revisions** are intended to incorporate enhanced methodologies resulting from updates to statistical methods,

concepts, data sources, definitions, or classifications used in collecting or compiling the affected data. These revisions often impact extended time series to minimise disruptions in continuity. Additionally, benchmark revisions frequently span multiple statistical domains and are carried out in a coordinated manner to maintain consistency across the affected areas.

20.2.2 Differentiation by Domain

Different domains have specific characteristics related to data and regulatory requirements from existing legal acts. Each domain has its revision policies, which vary in frequency, timing, length, depth, and reasons for revisions.

- **STBS** are routinely revised due to the continuous inflow of additional information and frequent benchmarking. Seasonal adjustments also contribute to these revisions. Monthly or quarterly STBS indicators are typically published within one to three months after the reference period ends. As a result, early estimates are made and revised later when more data becomes available. The weights used in **STBS** are based on **SBS**, which are released at least one year after the reference period. As a result, STBS weights need to be updated. Changes to the base year also affect the weighting system and lead to revisions.
- In **International Trade Statistics**, revisions to monthly and annual data may occur for up to two years after the reference period. News releases are based on a specific cut-off date, making them subject to revision. Traders may provide new information monthly, further influencing these updates.
- In the **Balance of Payments**, revisions are conducted quarterly, with previous data remaining open to revision. These revisions often incorporate results from newly published financial statements covering specific activities.
- Similarly, new information from financial statements is a common source of revision in National Accounts. A significant revision occurred with implementing the ESA 2010 methodology, including benchmarking to the year 2010. Following the adoption of ESA 2010, all National Accounts aggregates were revised for the entire time series up to 1995. Data from 1995 to 2010 are considered final, while GDP figures from 2011 onwards are provisional and subject to revision.
- In **government finance statistics**, debt figures for the three months preceding the reference period are considered provisional.
- **Labour cost indices** are revised quarterly to reflect updated data provided by primary sources.
- Seasonally adjusted data may need revisions after amending previously identified outliers. Seasonally adjusted data may occasionally be revised even if the unadjusted data remains unchanged.

20.3 Recommendations

20.3.1 Regular Revisions and Transparency

- Revisions should follow a regular pattern to avoid introducing bias.
- Data changes should only be made when absolutely necessary. Excessive revisions can create uncertainty.
- A balance must be struck between the value of additional information gained through revisions and the burden placed on users.
- The choice of data sources for statistical products, whether surveys or administrative records, should be based on professional considerations. Revisions must remain independent of political interference.
- A pre-announced revisions calendar should be in place to help users incorporate revisions into their analyses. Cycles of revision should be harmonised across statistical domains. Coordinating the dates and frequencies of routine revisions improves comparability and makes interpreting statistics easier.
- Significant errors should be corrected as soon as possible through timely revisions.
- The timing of revisions for seasonally adjusted and unadjusted data should be coordinated. Seasonally adjusted data should not be revised more frequently than unadjusted data.

20.3.2 Revision Policy

A clear and concise data revision policy should be in place for all published statistics.

- The policy should instil user confidence by providing advance explanations for revisions.
- The revisions policy should offer a general framework and be publicly accessible (see [01] References). It should be clear and consistent enough for users to anticipate future revisions.

20.3.3 Documentation

- The main reason for any revision must be clearly indicated, especially if the revision occurs on an ad-hoc basis. This helps reduce the risk of user misinterpretation.
- Changes should be explained in the relevant metadata.
- Documentation on methods for incorporating new data sources, data adjustments, and the frequency and size of revisions is essential.
- Users must be informed in advance about the potential revision of data, especially for preliminary estimates that are likely to be revised. Routine revisions should also be communicated ahead of time.
- Benchmark revisions should only be carried out when absolutely necessary. These revisions should be coordinated across statistical domains to prevent breaks in data. They must be pre-announced on the NSO website with a detailed description of the revision process.
- Data subject to revision and data that have been revised must be clearly flagged. Provisional figures should be marked with (p) to alert users that revisions may occur, allowing them to assess whether the data suits their purposes.
- The period during which figures are considered provisional must be clearly documented.
- For established statistical data series, users should be informed about the frequency and size of past revisions.
- A revisions schedule must be in place for domains that conduct revisions. This schedule should be publicly available and easily accessible.

20.3.4 Revision Analyses

- Revision analyses are essential to ensure data quality and to review revision procedures and methods.
- The mean absolute revision between provisional and final estimates is a useful indicator of the reliability of provisional statistics.
- This method is not recommended for annual data.
- The estimate should be based on at least 20 previous observations for quarterly indicators.
- The indicator is calculated as: $\frac{1}{n} \sum_{i=1}^n |l_i - p_i|$, where:
 - n = number of previous observations
 - l_i = late estimate
 - p_i = preliminary estimate

20.4 References

01. NSO's Policy on Revisions of Official Statistics: https://nso.gov.mt/wp-content/uploads/CDU-POL-002_Revisions-of-Official-Statistics-Policy.pdf
02. Eurostat European Commission. ESS Guidelines on Revision Policy for Principal European Economic Indicators. Available: <https://circabc.europa.eu/sd/a/37d2ae19-d15f-4cff-8b8b-14f39bca7fac/ESS%20guidelines%20on%20revision%20policy.pdf>
03. Eurostat European Commission. (2015). Short-term business statistics - revisions. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/Short-term_business_statistics_-_revisions

04. Eurostat European Commission. (2013). ESS Guidelines on Revision Policy. Available: <http://ec.europa.eu/eurostat/documents/3859598/5935517/KS-RA-13-016-EN.PDF/42d365e5-8a65-42f4-bc0b-aacb02c93cf7>.
05. Federal Statistical Office of Germany. (2016). General Revisions Policy. Available: https://www.destatis.de/EN/Methods/Quality/PrinciplesGoverningRevisions_of_Published.pdf?__blob=publicationFile.



21. Metadata Documentation and Quality Reporting

21.1 Scope and Purpose

Metadata can be defined as ‘data about data’. Indeed, statistical metadata has the role of describing and documenting statistical data. It reduces the complexities associated with statistical production, thus facilitating users’ understanding of statistical data. Altogether, statistical metadata promotes the appropriate use and sharing of statistical data over time. Metadata has a crucial role in the statistical production system. It provides enough information for users to discover, understand and use the disseminated statistical products. It is a way of informing users about how statistical data is compiled. Metadata enables broad re-use of information and establishes a knowledge base based on a time series of metadata documentation.

Compiling metadata in a statistical institution protects its reputation and reduces costs. It prevents the misuse of statistical products and limits unnecessary information requests from users.

Metadata documentation includes information on statistical concepts, definitions, standard classifications, code lists, collection methods, and data analysis techniques. It also supports reporting on key quality indicators. The information provided must be well-defined, relevant to users, reliable, timely, and easily accessible.

Documenting the statistical production process, including its phases and sub-processes, is essential. This allows for comparisons over time within the same domain and across different domains. Cross-domain comparisons help identify shared methods and systems while highlighting key methodological differences.

Metadata serves both statistical data users and producers, who often seek different types of information. Adequate metadata documentation provides a clear and logical description of statistical activities, covering both quality aspects and the entire process. This improves communication with users and between producers. The growing demand for statistics and analysis supports the need to make metadata widely available. Additionally, metadata documentation enhances transparency in statistical production.

There are three main types of metadata, namely Reference, Structural and Process-related.

21.2 Principles

21.2.1 Reference Metadata

Reference metadata describes the content and quality of statistical data, whether microdata or macrodata. A good report provides information on the main quality characteristics of a product, allowing users to assess its quality. It includes:

- The concepts used and how they function in practice (conceptual metadata).

- The methods used to produce the data (methodological metadata).
- The quality dimensions of the statistics produced (quality metadata).

The standard reporting structure for reference metadata is the Euro-SDMX Metadata Structure (ESMS), which replaced the Statistical Data Dissemination Standard (SDDS). ESMS documents methodological aspects, quality indicators, and statistical production processes. The current ESMS version includes 19 high-level concepts with limited sub-item breakdowns. While ESMS covers some quality concepts, a more detailed structure for quality indicator collection exists— the ESS Standard for Quality Reports Structure (ESQRS).

The current version of the **ESQRS** model includes 12 high-level concepts with various sub-item breakdowns.

The **ESMS** and **ESQRS** reporting structures are part of the **Single Integrated Metadata Structure (SIMS)**, which provides a comprehensive inventory of quality and metadata statistical concepts.

To address the needs of both data users and producers, a distinction is made between **user-oriented (U)** and **producer-oriented (P)** reporting structures. User-oriented reports focus on the general concepts covered by ESMS, while producer-oriented reports are more detailed and technical, aligning closely with the ESQRS structure.

Reference Metadata pertaining to the NSO may be accessible through the dedicated NSO Metadata Website: <https://metadata.nso.gov.mt/>.

21.2.2 Structural Metadata

Structural metadata refers to the set of metadata that provides information about the structure, organisation, and relationships of data within a dataset. Such metadata must be associated with a dataset to be meaningful. It describes how data is organised and presented, facilitating the correct interpretation and manipulation of statistical data. Structural metadata includes details such as:

- **Identifiers and Codes:** Information about unique identifiers and standardised codes (also known as codelists) used for variables, classifications, and categories in a dataset.
- **Relationships and Hierarchies:** Definitions of the relationships between variables or the hierarchical organisation of data categories (e.g., geographical or sectoral classifications).
- **Formats and Data Types:** Specifications of the data format, such as numeric, text, or date formats, and their constraints.
- **Dimensions and Measures:** Explanation of dimensions (e.g., time, geography) and measures (e.g., units of measurement, statistical indicators) that make up the dataset.
- **Geographical and Time References:** Information on the spatial and temporal coverage of the data.

Structural metadata is essential for ensuring that users can accurately understand, process, and analyse statistical data. It is also critical for interoperability, enabling datasets to be integrated, compared, or reused effectively across different systems or domains.

21.2.3 Process Metadata

Process metadata provides information on the production of statistics, from the initial identification of needs to the final stages of dissemination and evaluation. It covers the planning of processes, data collection, analysis, production, and dissemination.

Process metadata is normally retained for internal administrative purposes.

Ensuring quality is essential at every stage, and process metadata explains how quality is managed and safeguarded. It also details the decisions made, their reasoning, and how they will inform future developments in statistics. The **Generic Statistical Business Process Model (GSBPM)** is the standard for documenting process metadata. This model applies to all activities involved in producing official statistics, national or international. It can describe and assess the quality of processes based on administrative records, surveys, censuses, and mixed sources.

The GSBPM consists of eight phases, each with corresponding sub-processes. It is not a linear model, meaning sub-processes need not be followed in strict order. Some sub-processes can be skipped, while others may be revisited multiple times, creating iterative loops.

21.3 Recommendations

- Metadata documentation should be an integral part of the statistical production process across the organisation, ensuring better output.
- A metadata or quality report must accompany each statistical output for a specific domain. This report should concisely assess the content's reliability, quality, and applicability for various purposes.
- Documentation must cover concepts, definitions, classifications, standard code lists, methodology, quality aspects, and processes. It should be clear, complete, accurate, and transparent.
- Reports should follow a standard format. Metadata reports should adhere to the ESMS structure, quality reports to the ESQRS structure, and process reports to the GSBPM model.
- Reports should be tailored to the target audience. Reference metadata reports are aimed at data users, while process and quality reports are designed for data producers, future developers, and new colleagues. The level of detail and technicality should be adjusted to suit the audience.
- Metadata must be readily available and easily usable by internal and external users.
- Quality and methodological documentation should highlight the strengths and limitations of the data, allowing users to make informed decisions on how to use it.
- Providing metadata is crucial when major methodological revisions occur, with clear explanations of the reasons for the revisions and their impact.
- Metadata should be prepared promptly after completing the statistical activity. Delays in making metadata available can affect timeliness, relevance, and accuracy.
- Variations in metadata presentation standards should be minimised. Any changes must be documented and explained.
- A statistical metadata repository must store all metadata related to every statistical output. A structured storage system will make it easier for colleagues to access past documentation. Metadata should be organised by date and topic and entered and updated in a single location to ensure coherence.
- Relevant information must be requested from suppliers of administrative data used to produce statistics. This metadata allows the statistical institution to assess the suitability of the administrative data and take appropriate action. The information provided should meet the requirements of the institution's metadata template.
- Metadata must be updated regularly, with an internal framework to ensure compliance with update requirements. Generally, reference metadata should be updated annually (or less frequently for outputs produced less often), while process metadata should be updated every three years.
- Metadata reports must clearly indicate ownership, the date, and approval status.
- Metadata should be stored and disseminated in human-readable and machine-readable formats. The Statistical Data and Metadata eXchange (SDMX) is a standard exchange language that facilitates the sharing of statistical data and metadata among international organisations and member countries. Dissemination in SDMX format is possible when the ESMS and ESQRS standard structures are followed.
- Reference manuals and guidelines used for statistical projects should be stored for future reference.

21.4 References

Dippo, C. and Sundgren, B. (2000). 'The Role of Metadata in Statistics'. Proceedings of the Second International Conference on Establishment Surveys, June 17-21, 2000, Buffalo, New York. Available at: [program \(amstat.org\)](http://program.amstat.org) (Accessed 21 September 2022).

Eurostat (2014). *Technical Manual of the Single Integrated Metadata Structure (SIMS)* (semi-publication).

Gillman, D.W. and Appel, M.V. (1999). 'Statistical Metadata Research at the Census Bureau'. Annual Federal Committee on Statistical Methodology (FCSM) Research Conference, November 15-17. Available at: [Corel Office Document](#) (Accessed 21 September 2022).

United Nations Economic Commission for Europe (2009). *Statistical Metadata in a Corporate Context: A Guide for Managers*. Geneva: UNECE.

21.5 Other Sources

Eurostat's Concepts and Definitions Database: RAMON: 'Reference Metadata'. Available at: [Europa - RAMON - Nomenclature Detail View](#) (Accessed 21 September 2022).

Workshop on Statistical Metadata: Implementing the GSBPM and Combining Metadata Standards. Available at: [Workshop on Statistical Metadata: Implementing the GSBPM and Combining Metadata Standards | UNECE](#) (Accessed 21 September 2022).

Statistics New Zealand. Requirements for Metadata. Available at: [Requirements for metadata – Archives New Zealand](#) (Accessed 21 September 2022).



22. User Satisfaction Survey

22.1 Scope and Purpose

Evaluating whether users' needs are being met is crucial. This aligns with the NSO's responsibility to monitor product quality and analyse the use of statistics regularly. As a result, user satisfaction with statistical output must be regularly assessed. Users should be identified, and their statistical needs documented.

Statistical outputs are evaluated for quality and timeliness and their ability to meet users' needs. User satisfaction surveys are conducted to measure how well official statistics meet user demands. These surveys ensure that statistical information is useful to a broad range of users and aligned with current and potential needs. Additionally, user satisfaction surveys help identify the strengths and weaknesses of disseminated information and highlight areas most needing development and improvement.

The NSO keeps records of the number of news releases and publications disseminated on its website, the users who access statistical outputs, and the number of requests processed each year. Beyond assessing relevance, it is important to gauge users' perceptions of various statistical outputs. Satisfaction surveys are useful for monitoring the use of statistics and examining user opinions on these outputs.

User satisfaction surveys also help determine how official statistics are used for informed decision-making in government, business, education, research, and the media. The survey results can be used to improve the statistics' quality and highlight their potential. Additionally, solutions to increase public awareness and the importance of statistics in society can be explored.

With ongoing advances and evolving user needs, the statistics produced must keep up with user demands. For this reason, user satisfaction surveys are conducted periodically to monitor changes in statistical demand over time.

22.2 Principles

A user satisfaction survey gathers users' perceptions regarding the quality and timeliness of data produced. Users are asked about their views on the quality, timeliness, and usefulness of news releases, publications, and tailor-made requests. The survey also collects information on how NSO data is used and how its quality compares to statistics produced by other organisations. Additionally, the survey measures users' trust in NSO statistics and how frequently they request or access the NSO website.

Since users come from diverse backgrounds, analysis should account for these differences. The needs and demands vary between user groups, and gathering input from various users and purposes provides optimal results. This helps identify emerging needs and priorities.

22.3 Recommendations

22.3.1 General Considerations

- The anonymity of data providers must be safeguarded.
- User satisfaction surveys should ideally be conducted every two years.
- The target population should include government departments, entities, researchers, the media, international organisations, and other relevant user groups.
- The list of respondents should include all users who made tailor-made requests in the previous two years, as well as users subscribed to receive NSO news releases and other statistical products.
- Statistical offices should produce proxy measures without a user satisfaction survey, such as the number of questions received and complaints.
- The results of a user satisfaction survey should be made public, with accompanying documentation on lessons learned.

22.3.2 Questionnaire Content

- Information about improvement actions implemented due to the previous user satisfaction survey should be included. This can be presented in the introductory text accompanying the questionnaire.
- All statistical areas must be assessed.
- The questionnaire should include specific questions on tailor-made requests, access to main dissemination products, the NSO website, microdata access, and metadata access.
- Questions on the purpose of the requested information and the quality of support provided by NSO staff should also be included.
- The questionnaire should allow for the classification of users by type.

22.4 References

01. Agency for Statistics of Bosnia and Herzegovina (2012). *Quality report for statistical surveys methodological guidelines for preparation*. Sarajevo, Bosnia and Herzegovina: Agency for Statistics of Bosnia and Herzegovina.
02. Central Statistics Office (2006). *Standards and Guidelines, Quality is Statistics*. Cork, Ireland: Central Statistics Office.
03. Ecochard, P. et. Al. (2008). *User Satisfaction Surveys in Eurostat and the European Statistical System*. Rome, Italy
04. European Commission, Eurostat, (2016). *Follow-up of suggestions for improvements coming from the Eurostat user satisfaction survey*. Luxembourg: Eurostat European Commission.
05. Leadership Group (LEG) on Quality - Implementation Group, Eurostat European Commission (2003). *State-of-the-art regarding planning and carrying out Customer/User Satisfaction Surveys in NSIs*. Luxembourg: Eurostat European Commission.
06. Quality Unit, National Statistics Institute (INE) (2013). *User Satisfaction Survey for users of INE statistics*. Spain, Madrid: National Statistics Institute (INE).
07. Statistics Sweden, (2006). *Measuring Customer Satisfaction, a methodological guidance*. Sweden: Statistics Sweden.



23. Staff Satisfaction Survey

23.1 Scope and Purpose

In addition to providing independent, high-quality information, statistical institutes must ensure employees are satisfied with their jobs. Staff satisfaction drives continuous improvement and fosters trust in leadership, which are important for producing high-quality statistical output.

Staff satisfaction surveys should be conducted regularly, recognising that full staff engagement is vital for implementing a quality management system. These surveys enable the NSO to capture specific employee perceptions and opinions that might not surface during performance reviews or meetings. They allow employees to voice concerns anonymously, empowering them to contribute to workplace improvements regardless of their role. This can foster a positive attitude and make staff feel more valued.

Surveys also enhance trust and confidence in management, showing that leadership is open to listening and collecting valuable employee feedback regularly.

23.2 Principles

A staff satisfaction survey is a tool institutions use to gauge employee satisfaction. It assesses employees' feelings about leadership, the general work environment, workload, training, co-workers, job satisfaction, company policies, and work hours. Management can then address the main findings to improve planning and productivity.

Feedback from staff and inquiries by management are crucial for implementing targeted improvement measures. Staff satisfaction surveys are therefore essential for gaining genuine insights into management practices and employee satisfaction.

23.3 Recommendations

23.3.1 General Considerations

- Staff satisfaction surveys should be conducted every two years.
- Employees should be notified in advance when a staff satisfaction survey is being launched. The importance and scope of the survey should be clearly explained.
- Respondent anonymity must be protected, with the web-based collection being the preferred method. If employees require it, paper questionnaires should be made available. These anonymous forms should be safely returned via a suggestion box and then collected for data entry and processing.
- Questions that could identify a specific employee (e.g., job title) should be avoided or be non-mandatory.

- Survey results should be shared with both management and staff. The findings should inform an action plan that involves employee input. Staff should be updated on how the survey results will influence plans and decisions.
- A high response rate is desired, so the survey should be launched when employees are not facing tight deadlines (e.g., not at the end of a calendar quarter). The survey should allow for phased completion, enabling employees to start and finish later if necessary.

23.3.2 Questionnaire Content

- Information on improvement action plans adopted after the previous survey should be provided for follow-up surveys.
- Questions should cover job satisfaction, the image of the NSO, workload, training, teamwork, direct supervisors, perceived equality, remuneration, work-life balance, and the workplace environment. Likert scales should be used to measure levels of agreement.
- Open-ended questions should be included to allow employees to comment, as these responses often offer valuable feedback. These questions should be placed towards the end of the survey, just before the demographic section.
- Questions about demographic characteristics should be optional and placed at the end of the survey.

23.4 References

01. Hr-Survey. (2017). *What is an Employee Attitude Survey?* Available: <http://www.hr-survey.com/EmployeeAttitude.htm>. Last accessed 18th Sep 2017.
02. Kelleher, B (2011). *Dos and Don'ts for Conducting an Employee Engagement Survey*. Northeast Human Resources Association. Concord, MA, United States.
03. Markelevičius, J et al. (2009). *Quality Assurance Framework at Statistics Lithuania*. Lithuania: Statistics Lithuania.
04. Saari, L M et al. (2004). *Employee Attitude and Job Satisfaction*. Armonk, NY, United States: IBM Corporation, Global Workforce Research.

