

## Outbound Tourism Statistics (2019)

### NATIONAL REFERENCE METADATA IN SINGLE INTEGRATED METADATA STRUCTURE (SIMS)

#### **Concept 1 – Contact**

##### **Sub-Concept 1.1: Contact organisation**

National Statistics Office (NSO)

##### **Sub-Concept 1.2: Contact organisation unit**

Tourism and Education Statistics Unit

##### **Sub-Concept 1.3: Contact name**

Ms. Lara Pace

##### **Sub-Concept 1.4: Contact person function**

Principal Statistician

##### **Sub-Concept 1.5: Contact mail address**

National Statistics Office (NSO),  
Tourism and Education Statistics,  
Lascaris,  
Valletta, VLT 2000,  
Malta.

##### **Sub-Concept 1.6: Contact e-mail address**

lara.e.pace@gov.mt

##### **Sub-Concept 1.7: Contact phone number**

+356 2599 7630

#### **Concept 2 – Metadata update**

##### **Sub-Concept 2.1: Metadata last certified**

25th February 2020.

##### **Sub-Concept 2.2: Metadata last posted**

24th September 2020.

### Sub-Concept 2.3: Metadata last update

24th September 2020.

## **Concept 3 – Statistical presentation**

### **Sub-Concept 3.1: Data description**

The main aim of the Tourstat survey is to collect information on tourism demand, for both inbound and outbound trips with at least one overnight stay outside the usual environment. Tourism demand statistics include information on participation in tourism and information on tourism trips and visitors making the trips or same-day visits.

Tourism statistics are compiled primarily to present an overview of the volume of tourists, their characteristics and profile, total nights spent and tourism expenditure. This compilation of data helps public authorities and economic operators develop more suitable strategies and policies for improving the seasonal spread of holidays and tourism activities. The Tourstat survey is conducted to satisfy mainly the national needs.

### **Sub-Concept 3.2: Classification system**

The following statistical classifications are used for the Tourstat survey as a model instrument for coding variables and presenting data in an ordered manner:

- List of Countries - [ISO 3166](#), which is the International Standard for country codes that are used internationally.
- List of Airports.
- List of Accommodation Establishments - [NACE Rev. 2](#), which is the classification of economic activities corresponding to ISIC Rev. 4 (International Standard Industrial Classification of All Economic Activities) at European level.
- List of Localities - The local administrative units, abbreviated as [LAUs](#), form a system for dividing up the economic territory of the European Union (EU) for the purpose of statistics at local level based on the geographical location of units. They have been set up by Eurostat and they are compatible with [NUTS](#) (Nomenclature of Territorial Units for Statistics). At local level, two levels of LAU have been defined: The upper level (LAU1, formerly NUTS level 4) is defined for most, but not all, of the countries. The lower level (LAU2, formerly NUTS level 5) consists of municipalities or equivalent units in the EU Member States.
- List of foreign currencies.

### **Sub-Concept 3.3: Sector coverage**

All those going through departures at either the airport or seaport during an interviewing period may be invited to take part in the survey. This description excludes:

- Employees on duty: Including airline/ shipping crew and MIA/Seaport personnel.
- Passengers and crew on domestic flights.

- Re-crosses – i.e. Passengers who have already crossed the counting line on the sample day even if it was before the shift.
- Cancelled flights – If a flight is cancelled and the team leader can find out from desk staff that all the passengers have been re-booked airside onto another flight/sailing (i.e. they would not come through the line again), the interviews should be left as eligible. If the passengers will leave the airport/seaport and will have to pass through passport control again, then all passengers on that flight/sailing are ineligible (re-crosses).
- Passengers are eligible if they are departing on a flight which has been delayed overnight
- Passengers who are refused admission and/or are being escorted through Immigration controls by security or police officers are ineligible.

### Sub-Concept 3.4: Statistical concepts and definitions

The statistical variables that are measured via the Tourstat questionnaire are as follows:

- Usual environment: The geographical area within which an individual conducts his regular life routines and shall be determined on the basis of the following criteria: the crossing of administrative borders or the distance from the place of usual residence, the duration of the visit, the frequency of the visit, the purpose of the visit.
- Traveller: A person who moves between different geographic locations, for any purpose and any duration.
- Visitor: The three fundamental criteria used to distinguish visitors from travellers are:
  - a. The trip should be to a place other than that of the usual environment, which would exclude short-distance local transport and commuting, i.e. more or less regular trips between the place of work/study and the place of residence;
  - b. The stay in the place visited should not last more than twelve consecutive months, beyond which the visitor would become a resident of that place (from the statistical standpoint);
  - c. The main purpose of the visit should be other than the exercise of an activity remunerated from within the place visited, which would exclude migratory movements for work purposes.
- Same-day visitor: A visitor who does not spend the night in collective or private accommodation in the place/country visited.
- Tourist: A visitor who stays at least one night in collective or private accommodation in the place/country visited.
- Tourism: The activity of visitors taking a trip to a main destination outside the usual environment, for less than a year, for any purpose, including business, leisure or other personal purpose, other than to be employed by a resident entity in the place visited. A person who is employed within the country visited is not defined as a tourist (i.e. can't be classified as a tourist travelling for business purposes).
- Inbound tourism: The activities of non-resident visitors travelling to Malta (i.e. outside their usual environment) and staying for not more than 12 consecutive months for leisure, business or other (corresponding) purposes.
- Outbound tourism: the activities of residents of Malta travelling to and staying in places outside Malta and therefore outside their usual environment.
- Nationality Passport - The country passport that the passenger is using.
- Country of Residence – The country a person has been living (studying or working) in for the past 12 months or the intention is to stay in the country for 12 months or more.
- Airport/Seaport embarked from to visit Malta - The airport/seaport must be the one the respondent embarked from for their journey to Malta. Thus, if the interviewee had to catch a connection flight on his way to Malta, only the final leg of the journey should be recorded.

- Airline Company used for the outgoing journey - The airline company the respondent is choosing to travel with.
- Airport/Seaport going to on this trip - The airport in which the respondents are going to from Malta.
- End/Change Airport - If an interviewee is a foreigner spending more than one night in a country after leaving Malta, he/she should be marked as his/her final destination (End). Foreigners forced to spend one night in an airport/hotel due to flight connectivity, should be marked as transit passengers (Change).
- Final destination – The place where the trip from Malta will end. For outbound tourists, it is defined as the place visited that is central to the decision to take the trip. If no such place can be identified by the visitor, the main destination is defined as the place where the respondent spent most of the time during the trip.
- Nights spent abroad – Duration of the trip in number of nights. This variable is also used for estimation of migration components.
- End of visit/Will return to continue visit - A respondent is classified as a returning passenger if: While he/she is in Malta, decided to go for a few days abroad and will be returning to continue his/her visit in Malta. The main aim of this question is to avoid double counting of departing passengers.
- Gozo and/or Comino visit – To estimate how many tourists visited Gozo and/or Comino while they were in Malta.
- If the respondent spent at least one night in Malta – This question is intended to filter tourists from same day visitors.
- Main purpose of the trip –This is defined as the purpose for which the trip was taken. In absence of this main purpose, the trip would not have been taken. The main purposes which are asked in the Tourstat survey are as follows:
  - Holiday/Leisure, boarding a cruise ship, visiting relatives and friends
  - Professional/business, travelling on an incentive trip, attending a conference or large business meeting, paid employed from within the country visited, a member of embassy personnel
  - Studying English at a language school, health treatment, other purposes
- Frequency of visit - Determines the number of times a respondent visited Malta:
  - Less often than once a year
  - At least once a year
  - 2 or 3 times a year
  - More than 3 times a year but less than once a week
  - At least once a week
- Number of nights spent in Malta (including Gozo/Comino) - Duration of the inbound trip in number of nights, including nights spent in Gozo and/or Comino.
- Locality stayed longest – In this question only the locality where the visitor stayed longest should be noted.
- Modes of transport – This is a multiple question which asks the visitors to indicate the different modes of transport used to get around Malta/Gozo: Bus, Taxi, Coach, Private vehicle, Car/motorcycle hire
- Package trip – A pre-arranged combination of not fewer than two of the following when sold or offered for sale at an inclusive price and when the service covers a period of more than twenty-four hours or includes overnight accommodation:
  - Transport
  - Accommodation
  - Other tourist services not ancillary to transport or accommodation and accounting for a significant proportion of the package.

- Booking of the trip – Use of tour operator or travel agency to book the main means of transport and/or accommodation. Alternatively, main means of transport and/or accommodation can be booked via internet.
- Timeshare accommodation - This question aims to determine the use of timeshare versus the use of other types of accommodation arrangements. Time share involves the payment of a fee by which a person can use any timeshare accommodation owned by the company.
- Main means of accommodation - An interviewer is meant to ask respondents what type of accommodation they made use of while in Malta/ intend to make use of while abroad from one of the following options:
  - Hotel
  - Guesthouse
  - Aparthotel
  - Hostel
  - Other collective accommodation
  - Holiday furnished premises (including farmhouses)
  - Other rented accommodation (including cruise ship)
  - Own private residence
  - Staying with friends or relatives
  - Other private accommodation (including time share)
- Hotel category - This question applies only to those respondents which made use of hotel accommodation to determine the star category.
- Type of accommodation service – Room only, Bed and Breakfast, Half Board, Full board or All inclusive.
- Total Expenditure – Maltese residents are asked to budget the expenditures which they are about to incur when travelling abroad, while foreign residents are asked for an account of the expenses incurred in travelling to Malta. The total consumption expenditure is broken down by the following expenditure items:
  - Package Expenditure - This cost includes transport and accommodation services as well as travel agent commissions and other administrative charges. This variable is subdivided into 3 parts:
    - 1) Type of currency
    - 2) Amount of the package expenditure
    - 3) Number of persons covered
  - Non-Package Expenditure - Air/Sea Fare and Accommodation expenditure. This variable is subdivided into 3 parts:
    - 1) Type of currency
    - 2) Amount of the travel and accommodation expenditures
    - 3) Number of persons covered
  - Other Expenditure - Sum up all other expenditures incurred during their stay in Malta/ visit abroad. Examples include shopping, souvenirs, tickets for concerts or a sports event, entrance to a museum or zoo, day excursions during a tourism trip. This variable is subdivided into 3 parts:
    - 1) Type of currency
    - 2) Amount of other expenditure
    - 3) Number of persons covered
- Age group – 0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65+.
- Gender – Male or Female.

### **Sub-Concept 3.5: Statistical unit**

- Tourism trips with overnights stays: The statistical unit is the trip with at least one overnight stay made by the individual.
- Same-day visits: The statistical unit is the same-day trip made by the individual.

### **Sub-Concept 3.6: Statistical population**

Inbound tourism: The activities of a non-resident visitor within Malta on an inbound trip.

Outbound tourism: The activities of a resident visitor outside Malta as part of an outbound trip.

### **Sub-Concept 3.7: Reference area**

All departing passengers at the Malta International Airport.

### **Sub-Concept 3.8: Time coverage**

From 2001 onwards.

### **Sub-Concept 3.9: Base period**

Not applicable.

### **Concept 4 – Unit of measure**

#### Tourism trips:

The respondent selected for the interview will report on the overnight trip(s) made during the reference period and give the details for the different characteristics of the trip and his/her socio-demographic profile (the visitor making the trip). Each observed trip is an individual record in the micro-data. Data for expenditure on tourism trips is expressed in Euros.

### **Concept 5 – Reference period**

2019.

### **Concept 6 – Institutional mandate**

#### **Sub-Concept 6.1: Legal acts and other agreements**

[The Malta Statistics Authority Act, 2000.](#) The Act empowers the NSO to collect, compile, extract and release official statistics related to demographic, social, environment, economic and general activities and conditions of Malta.

#### **Sub-Concept 6.2: Data sharing**

Apart from a quarterly and monthly news releases, users may also request data both in aggregate format as well as in the form of anonymised micro data.

Users can request information through the NSO website:

- <https://nso.gov.mt/en/Services/Pages/Request-for-Information.aspx>
- <https://nso.gov.mt/en/Services/Microdata/Pages/Access-to-Microdata.aspx>

The tourism statistics are published mainly in the following dissemination products:

- [Monthly Inbound Tourism](#) news release;
- [Quarterly Outbound Tourism](#) news release;
- [Yearly Regional Tourism](#) news release;
- [Malta in Figures](#) publication;
- [Gozo in Figures](#) publication;
- Economic Survey requested by the Ministry for Finance.

## **Concept 7 – Confidentiality**

### **Sub-Concept 7.1: Confidentiality – Policy**

At National level:

The NSO requests information for the compilation of official statistics according to the articles of the MSA Act – Cap. 422 and the Data Protection Act – Cap. 586 of the Laws of Malta implementing the General Data Protection Regulations (GDPR).

Article 40 of the MSA Act stipulates the restrictions on the use of information while Article 41 stipulates the prohibition of disclosure of information. Furthermore, Section IX of the Act (Offences and Penalties) lays down the measures to be taken in case of unlawful exercise of any officer of statistics regarding confidentiality of data.

Since its inception, the NSO has always assured that all data collected remains confidential and that it is used for statistical purposes only according to the articles and derogations stipulated in the laws quoted above. The Office is obliged to protect the identify of data providers and refrain from divulging any data to third parties that might lead to the identification of persons or entities.

During 2009, the NSO has set up a Statistical Disclosure Committee to ensure that statistical confidentiality is observed, especially when requests for microdata are received.

Upon employment, all NSO employees are informed of the rules and duties pertaining to confidential information and its treatment. In line with stipulations of the MSA Act, before commencing work, every employee is required to take an oath of secrecy whose text is included in the same Act.

An internal policy on anonymisation and pseudo-anonymisation is in place to ascertain that adequate methods are used for the protection of data which the office collects and shares with the public in its capacity as the National Statistics Office. The policy is meant to safeguard confidentiality of both personal and business data entrusted to the NSO. The document provides guidance for all NSO employees who process data on a daily basis as to how anonymisation and pseudo-anonymisation methods should be applied. The policy applies to all confidential, restricted and internal information, regardless of form (paper or electronic documents, applications and databases) that is received, processed, stored and disseminated by the NSO.

At European level:

[Regulation \(EC\) No 223/2009](#) on European statistics (recital 24 and Article 20(4) of 11 March 2009 (OJ L 87, p. 164), stipulates the need to establish common principles and guidelines ensuring the confidentiality of data used for the production of European statistics and the access to those confidential data with due account for technical developments and the requirements of users in a democratic society.

### **Sub-Concept 7.2: Confidentiality – Data Treatment**

In order to guarantee confidentiality of personal data, only aggregated data are normally published (observation units are not recognizable either directly or indirectly). Furthermore, the rules below are followed to ensure that unreliable or confidential counts are not published:

- Estimation based on 20 to 49 sample observations => flag 'u' (unreliable)
- Estimation based on fewer than 20 observations => flag 'c' (confidential)

### **Concept 8 – Release Policy**

#### **Sub-Concept 8.1: Release Calendar**

An advance release calendar is maintained by the NSO and published on the NSO website. The calendar projects three months of news releases (including the current and two subsequent months).

#### **Sub-Concept 8.2: Release Calendar access**

[https://nso.gov.mt/en/News\\_Releases/Release\\_Calendar/Pages/News-Release-Calendar.aspx](https://nso.gov.mt/en/News_Releases/Release_Calendar/Pages/News-Release-Calendar.aspx)

#### **Sub-Concept 8.3: User access**

An internal policy on dissemination is in place to govern the dissemination of official statistics in an impartial, independent and timely manner, making them available simultaneously to all users.

The NSO's primary channel for the dissemination of official statistics is the NSO website. Tailored requests for statistical information may also be submitted through the NSO website.

Moreover, tourism statistics' news releases and publications are available in electronic format on the NSO website.

In addition, tabular information can be made available upon request.

The data are very sought by the media, researchers, students and policy makers. Such users can request information through the NSO website (<https://nso.gov.mt/en/Services/Pages/Request-for-Information.aspx>) where data are later provided at an aggregated level.



## **Concept 9 – Frequency of Dissemination**

The Tourstat survey results are published every month and on a quarterly basis.

On an international level, information on outbound same-day tourism statistics are sent to Eurostat on an annual basis, reporting separately the four quarters of the previous calendar year as outlined in sub-process 2.1. Aggregated data can be assessed on the following website:

<http://ec.europa.eu/eurostat/data/database>

- > Databases by themes
- > Industry, trade and services
- > Tourism (tour)

Inbound and outbound tourism statistics are available by gender and age group through the online statistical database (StatDB). This information is made available on a monthly basis.

## **Concept 10 – Accessibility and Clarity**

### **Sub-Concept 10.1: News release**

News releases are published on the NSO website and may be accessed through the following links:

- Inbound Tourism news release:

[https://nso.gov.mt/en/News\\_Releases/View\\_by\\_Unit/Unit\\_C3/Tourism\\_Statistics/Pages/Inbound-Tourism.aspx](https://nso.gov.mt/en/News_Releases/View_by_Unit/Unit_C3/Tourism_Statistics/Pages/Inbound-Tourism.aspx)

- Outbound Tourism news release:

[https://nso.gov.mt/en/News\\_Releases/View\\_by\\_Unit/Unit\\_C3/Tourism\\_Statistics/Pages/Outbound-Tourism.aspx](https://nso.gov.mt/en/News_Releases/View_by_Unit/Unit_C3/Tourism_Statistics/Pages/Outbound-Tourism.aspx)

- Regional Tourism news release:

[https://nso.gov.mt/en/News\\_Releases/View\\_by\\_Unit/Unit\\_C3/Tourism\\_Statistics/Pages/Regional-Tourism.aspx](https://nso.gov.mt/en/News_Releases/View_by_Unit/Unit_C3/Tourism_Statistics/Pages/Regional-Tourism.aspx)

### **Sub-Concept 10.2: Publications**

Tourism statistics are featured in the following publications:

- [Malta in Figures](#) publication.
- [Gozo in Figures](#) publication.

### **Sub-Concept 10.3: Online Database**

Inbound and outbound tourism statistics are available by gender and age group through the online statistical database ([StatDB](#)). This information is made available on a monthly basis.

#### **Sub-Concept 10.4: Micro-data access**

Anonymised microdata for Tourism statistics may be provided under strict conditions to a selected number of institutions or persons accredited as research entities or researchers respectively. Further information on access to anonymised microdata is available on the NSO website through: <http://nso.gov.mt/en/Services/Microdata/Pages/Access-to-Microdata.aspx>.

Researchers who require such access need to submit an application form clearly explaining the purpose of their statistical research and justifying their need for access to microdata. The application form will be evaluated internally and if considered favourably a formal contractual agreement will be drafted to explain the responsibilities of the researcher for the security of the information. Once the agreement is agreed upon and signed by both parties, access to anonymised microdata will be granted subject to the terms of reference included in the contractual agreement. Access is normally granted for a definite time period.

#### **Sub-Concept 10.5: Other**

Not applicable.

#### **Sub-Concept 10.6: Documentation on methodology**

Work processes and procedures for the compilation of Inbound/Outbound tourism statistics are documented in a standardised reporting template and aligned to the GSBPM model. The model covers all phases of the statistical production process, from the initial stages of identifying what statistics are needed and the scope of the particular survey, to the final stages of dissemination and evaluation. The GSBPM report is only available internally and may be accessed by all NSO employees.

In addition, methodological notes accompany each News Release being published.

#### **Sub-Concept 10.6.1: Metadata completeness rate**

Information about all required metadata concepts (and sub-concepts thereof) are provided.

#### **Sub-Concept 10.7: Quality Documentation**

As from 2018, no metadata related to Tourstat was being delivered to Eurostat. Tourstat is currently being used to calibrate outbound tourist trips in the National Tourism (NT) Survey (of which metadata report of NT is sent to Eurostat every September).

Tourism statistics SIMS reports are available to the public on the [NSO's metadata website including concepts related to metadata and quality](#).

The NSO has developed an internal Quality Management Framework (QMF) which is built on common requirements of the ESS Code of Practice (ESS CoP). A document was prepared to include a set of general quality guidelines spanning over all statistical domains. Assuring methodological soundness is an integral part of the QMF, nonetheless, the document spans also on other areas related to institutional aspects.

## **Concept 11 – Quality Management**

### **Sub-Concept 11.1: Quality Assurance**

The Tourstat survey is deemed to be a reliable survey. This is achieved through the use of standard guidelines which are made available to all interviewers.

The NSO ensures the accuracy of data released to the public and prepares clear methodological notes which explain the processes involved in the collection and production of official statistics.

The NSO has developed an internal Quality Management Framework (QMF) which is built on common requirements of the ESS Code of Practice (ESS CoP). A document was prepared to include a set of general quality guidelines spanning over all statistical domains. Assuring methodological soundness is an integral part of the QMF, nonetheless, the document spans also on other areas related to institutional aspects.

Every five to seven years, the NSO participates in a Peer Review exercise through which the compliance of its operations with principles of the ESS CoP is assessed by an expert team. Peer Reviews are indeed part of the European Statistical System (ESS) strategy to implement the ESS CoP. Each NSI is expected to provide information as requested by a standard self-assessment questionnaire. Following this an expert team visits the office to meet NSI representatives and main stakeholders. Peer Reviews result in a compliance report and the listing of a set of Improvement Actions which need to be followed up by the NSI. The next round of Peer Reviews is planned to be carried out in 2022.

### **Sub-Concept 11.2: Quality Assessment**

- For data collection, interviewers are trained and briefed by the NSO to ensure quality and comparability.
- Monitoring and supervising the interviewers allows detecting where there is room for improvement.
- The data are checked regularly during data collection, to improve data collection methods where necessary.
- Data treatment is clearly documented.
- A focus on the Quality Management Framework (QMF) for the NSO in order to ensure consistency with the European Statistics Code of Practice.
- The European Statistics Code of Practice sets the standard for developing, producing and disseminating European statistics - <https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>

## **Concept 12 – Relevance**

### **Sub-Concept 12.1: User needs**

Tourism statistics are compiled primarily to present an overview of the volume of tourism trips, characteristics and profile of the tourist, total nights spent and tourism expenditure. This compilation of data helps public authorities and economic operators develop more suitable strategies and policies for improving the seasonal spread of holidays and tourism activities. Tourstat survey is conducted to satisfy mainly the national needs.

Tourism statistics are widely used both internally (within NSO) and externally. Within the NSO, main users of the Tourstat data include the following statistical domain units:

- Regional Statistics
- National Accounts,
- Balance of Payments
- Price Statistics
- Population and Migration Statistics

On a national level, tourism statistics are used in policy making by Ministries, Government departments and Authorities working within the social and economic fields. These figures are also used by research institutions, the University of Malta, other educational institutions, students and private companies. The main external stakeholders who make use of the tourism statistics compiled include:

- [Malta Tourism Authority \(MTA\)](#);
- [Central Bank of Malta](#);
- [Ministry for Tourism and Consumer Protection](#);
- [Malta Hotels and Restaurants Association \(MHRA\)](#);
- [Deloitte](#);
- [PKF Malta](#);
- [Australia High Commission](#).

On an international level, the characteristics of outbound same-day visits are transmitted to Eurostat on an annual basis (reporting separately the four quarters of the previous calendar year) to satisfy the requirements laid down in [Regulation \(EU\) No 692/2011](#) (Annex 1) concerning European statistics on tourism and repealing [Council Directive 95/57/EC](#). The first reference period was on 1st January 2014.

### **Sub-Concept 12.2: User satisfaction**

The last User Satisfaction Survey was held in 2014 with the aim to collect information about key users' satisfaction with statistical output.

The NSO keeps record of the number of News Releases and publications disseminated on its website; the users to whom statistical products are provided; as well as the number of requests that are processed every year.

News Releases and tailor-made statistical outputs were assessed on account of their quality, timeliness, and on their ability to meet users' needs.

### **Sub-Concept 12.3: Data Completeness**

Not applicable. TOURSTAT survey is conducted to satisfy the national needs. It is neither regulated by the MSA Act 2000 nor by an EC Regulation.

## **Concept 13 – Accuracy and reliability**

### **Sub-Concept 13.1: Overall accuracy**

- The main possible source of error comes from the memory effect (where respondents might forget trips and details about their expenditure). This would lead to an underestimation of trips and expenses.

- Sampling errors vary among variables. The main variables are relatively accurate (e.g. Number of trips with purpose "leisure, recreation and holidays") whereas more detailed variables have a larger sampling error (e.g. expenditure related questions).
- Non-response errors: Significant item non-response appears for variables referring to expenditures. When dealing with item non-response, regression imputation methods are used.
- Measurement errors:
  - By interviewers: Arise due to the possibility of not understanding or misinterpreting the respondent's answers and typing errors
  - By respondents: Arise due to the possibility of having respondents not understanding the questions, not knowing the answers, not interested in the survey and an unwillingness to provide proper information on sensitive questions such as those on expenditure. The main possible source of error comes from recall bias (whereby a respondent might forget some details on trips, such as expenditure). This would lead to an underestimation of both trips and expenses.
- To avoid the publication of figures which are statistically unreliable, unreliable and under-represented figures are indicated for each table in the news releases and are flagged to be treated with caution.
- The NSO advises users to exercise caution when analysing unreliable estimates, which are based on less than 20 sample observations and under-represented estimates, which are based between 20 and 49 sample observations.

### Sub-Concept 13.2: Sampling errors

- Inbound tourists – 2,753,239 ± 15,102 (CV: 0.28%)
- Total nights spent by inbound tourists – 19,338,860 ± 224,541 (CV: 0.59%)
- Total expenditure by inbound tourists – 2,220,627 ± 17,480,582 (CV: 0.40%)
- Outbound tourists – 706,797 ± 10,987 (CV: 0.79%)
- Total nights spent by outbound tourists – 4,608,643 ± 81,565 (CV: 0.90%)
- Total expenditure by outbound tourists – 606,220 ± 9,535,973 (CV: 0.80%)

Note: Expenditure estimates are rounded to the nearest thousand.

- The bootstrap method is applied on micro data by means of SPSS in order to estimate the variance and sampling errors – coefficients of variation, for the number of outbound and inbound tourists, total nights and total expenditure. This helps in checking the reliability of the sample estimates.
- From the resulting readings of the bootstrap, the mean, standard error about the mean (SE), standard deviation and variance are computed. These results are then used to work the coefficient of variations (CV) and the 95% confidence interval of the estimated values.

#### Outlier detection techniques:

Two variables that are most likely to be affected with outliers are the number of nights spent and the expenditure categories. The standardized residuals, which are obtained by fitting a multiple linear regression model, are the most commonly used measures for detecting outliers. Observations with resulting standardized residuals exceeding +2 or -2, indicates that one should

investigate these data points. An alternative measure that is used for indicating data points that are particularly worth checking for validity is Cook's distance. A conventional cut-off value that is used for spotting highly influential points is when the resulting Cook's distance is greater than  $4/n$ , where  $n$  is the sample size. The two methods are used to detect outliers for the number of nights spent and for the different expenditure categories.

### **Sub-Concept 13.3: Non-sampling error**

Refer to sub concept 13.1 for main sources of non-sampling error.

### **Sub-Concept 13.3.1: Coverage error**

Not applicable.

### **Sub-Concept 13.3.1.1: Over Coverage**

Not applicable. Travellers are only counted once on route to the departures lounge.

### **Sub-Concept 13.3.1.2: Common Units Proportion**

Not applicable.

### **Sub-Concept 13.3.2: Measurement error**

Measurement errors by respondents: May arise due to the possibility of not understanding the questions, not knowing the answers (not remembering details), not being interested in the survey and an unwillingness to provide proper information on sensitive questions such as expenditure. Main possible source of error comes from recall bias (whereby a respondent might forget some details on trips, such as expenditure). This may lead to over- or underestimation of trips and expenses.

Measurement errors by interviewers: Maybe arise due to the possibility of not understanding or misinterpreting the respondent's answers, typing errors, recall bias and interviewers' bias (Questions laced with interviewer bias can influence respondents in such a way that distorts the outcome of the interview.)

In case of business trips organized by employers using services of travel agencies, respondents do not know the costs of transport and other services. Significant item non-response appears for variables referring to travel ticket/accommodation/other expenditures. For dealing with item non-response, regression imputation methods are used.

Quality improvements:

- New in-built validations in the data collection program;
- Improvement in the wording of the questions to minimize interviewers' bias especially in the type of accommodation question;
- Improvement in the length, design and complexity of the questionnaire.
- Intensive interviewer's training;
- New questions are introduced in the Tourstat survey so as to increase the checks in the quality of the main indicators for instance – the time-share accommodation was added in

the survey so as to check the quality of the main means of accommodation. Similarly, when the Libya crisis was ongoing, it was deemed important to deduct the number of persons that embarked from Libya and came to Malta, thus, an extra question that deals with embarkation airports was added.

### **Sub-Concept 13.3.3: Non-response error**

Out of a total of 88,357 sample passengers, 34,155 individuals (38.7%) were clicked/refusals/ non-contacts, such that:

- Number of clicked: 2,408 individuals (or 3.6%)
- Number of refusals: 29,178 individuals (or 30.6%)
- Number of non-contacts: 2,569 individuals (or 2.9%)

Methods used for dealing with unit non-response: Figures are then inflated by the departing passenger movements provided by the Malta International Airport.

Significant item non-response appears for variables referring to expenditures. For dealing with item non-response the regression imputation methods are used.

### **Sub-Concept 13.3.3.1: Unit non-response**

The unit non-response rate amounts to approximately 38.7% of total passengers in the sample. The response rate for each interviewer on a monthly basis is monitored by the Data Management Unit to improve their performance in the future. The main causes of unit non-response are: Respondents might not understand the questions, do not remember the details of their trip, are not interested in the survey or not willing to provide proper information on sensitive questions such as expenditure. Main possible source of error comes from recall bias (whereby a respondent might forget some details on trips, such as expenditure). This may lead to over- or underestimation of trips and expenses.

The following methods are used to deal with unit non-response:

- Correction factor in the weighting procedure (Extrapolation), Administrative records are used as a supplementary source for the calibration of the survey aggregates. The exercise is in the main bottom-up approach including a calibrating weight to re-align the survey results to the true population of total passengers' departures (net of transits).
- Partial or minimum responses are also accepted if these cover the main indicators about respondents collected in the questionnaire.

### **Sub-Concept 13.3.3.2: Item non-response**

Item non-response rates for the main variables are approximately as follows:

- Purpose of visit – 2.8%
- Organisation of stay (Package/Non-Package Trip) – 8.6%
- Frequency of visit – 4.0%
- Nights spent - 1.1%
- Package expenditure – 26.0%
- Travel ticket expenditure – 27.8%
- Accommodation expenditure – 27.3%

- Other expenditure – 23.7%

It was observed from past datasets that the type of accommodation variable consisted of a large amount of missing data (25-35%) and this was affecting the aggregate estimates. For this purpose, as from 2015 onwards, a survey is classified as partial if the question for type of accommodation is complete. This enables the interviewers to make more effort in obtaining more information on the type of accommodation the respondent (both if foreign or Maltese) used/will use. In fact, this made a significant impact in the imputation rates as the item non-response rate for the type of accommodation was lowered to less than 0.5%.

#### **Sub-Concept 13.3.4: Processing error**

In-built validations are available in the data collection program. The unit continues to do improvements on the Tourstat CAPI (Computer Assisted Personal Interviewing) and the controls for data entry in order to minimize errors and missing values for the core variables (i.e. expenditures and socio-demographic characteristics).

The main variables are relatively accurate (e.g. Number of trips with purpose "leisure, recreation and holidays") whereas more detailed variables have a larger sampling error (e.g. Total Expenses). Imputation is kept at a minimum for the core variables (purpose of visit, destination, type of accommodation used etc.) as interviewers are instructed to get complete answers. Most of the imputations are carried out on expenditure categories, which provides the unit with very accurate expenditure estimates. No proxy interviews are allowed unless the passenger selected is underage.

#### **Sub-Concept 13.3.5: Model assumption error**

Not applicable.

### **Concept 14 – Timeliness and punctuality**

#### **Sub-Concept 14.1: Timeliness**

##### Tourstat – Time Frames

<b>Activity</b>	<b>Time</b>
Data collection at the MIA departures	Daily
Filing and updating of exchange rates	Daily
Sending requests for administrative data from external sources	T+10 days
Updating the exchange rates in the database and sampling rates checks	T+15 days
Data validation	T+15 days
Data cleaning	T+20 days
Weighting procedure	T+25 days
Selecting inbound and outbound tourists from the total passenger (filtering out of non-tourists)	T+25 days
Preliminary Analysis (Comparing results with previous years)	T+25 days



Sampling errors	T+26 days
Dissemination – Monthly Inbound Tourism news release	T+30 days
Dissemination – Quarterly Outbound Tourism news release	T+45 days
Dissemination – Yearly Regional Tourism news release	T+6 months
Information on outbound same day visits are sent to Eurostat	T+6 months

### **Sub-Concept 14.2: Punctuality**

All releases are published and disseminated at 11:00 a.m. as scheduled in the Advance Release Calendar. Considering all tourism news releases disseminated during the year 2019, no news release was late.

## **Concept 15 – Coherence and Comparability**

### **Sub-Concept 15.1: Comparability – Geographical**

There have been no geographical comparability problems recorded.

Common tourism definitions are used between Malta and other EU member states following the Methodological Manual provided by Eurostat, however, methods used for data collection vary between countries.

### **Sub-Concept 15.2: Comparability**

A comparable time series is available from 2001 onwards.

### **Sub-Concept 15.3: Coherence – Cross Domain**

For the reasons outlined below, the data can be seen as comparable to other related statistical outputs.

- A representative from the MIA supplies this office on a monthly basis with MIA monthly passenger movements in order to be used for the weighting procedure of air passengers for calibration of the survey aggregates.
- The number of transit passengers is provided on a monthly basis by a representative from MIA in order to deduce departures for tourism system from the total MIA passenger departures.
- Information on the number of overnight cruise passengers is provided on a monthly basis by a representative from Transport Malta to include the figures in the Inbound Tourism news release.
- Administrative data records of tourist sea departures are provided by a representative of Virtu Ferries Ltd. in order to be used as a supplementary source for the weighting procedure of sea passengers. These include:

- Passengers departing Sicily by nationality and gender (Arrivals),
- Passengers departing Malta by nationality and by gender (Departures),
- Day Return Passengers from Sicily (Seats).

#### **Sub-Concept 15.3.1: Coherence – Sub-Annual and Annual statistics**

Not applicable.

#### **Sub-Concept 15.3.2: Coherence – National Accounts**

Not applicable.

#### **Sub-Concept 15.4: Coherence – Internal**

When comparing non-resident arrivals and nights from TOURSTAT and ACCOMSTAT, one will notice some discrepancies, mainly due to different definitions and methodologies. Some of these differences include:

- Two-centre holidays: For example, if a tourist spends 4 nights in one hotel and the remaining 3 nights of his stay in a different hotel, the tourist will be counted twice in ACCOMSTAT as data is collected directly from hotel registers. On the other hand, this tourist will be counted only once in TOURSTAT.
- Main type of accommodation used: TOURSTAT records the main type of accommodation used by the tourist during his stay. If for example, a tourist spends 4 nights in private accommodation and 3 nights in a 5-star hotel, all 7 nights will be assigned to private accommodation. On the other hand, in ACCOMSTAT, these nights will be split according to the hotel categories.
- Time-share accommodation: Time-share accommodation is classified as private accommodation in TOURSTAT. On the other hand, collective accommodation establishments furnish the NSO with a total figure, thus it is not possible to identify the number of arrivals and nights spent on time-share.
- Definition of a tourist: In TOURSTAT, one can distinguish between tourists and non-tourists, for example paid employed from within the country visited. The latter are excluded from TOURSTAT and present in ACCOMSTAT. Thus, in the case of foreign residents, arrivals in hotel accommodation do not necessarily comply with all the criteria defining a tourist.
- When a passenger is interviewed after midnight, TOURSTAT assigns that individual an extra night, even though the tourist didn't stay in the hotel in Malta for that night in question.
- Residents and non-residents in ACCOMSTAT are counted during their month of arrival, while in TOURSTAT they are measured on departure.
- Long stays (exceeding 30 nights) are treated differently in ACCOMSTAT and TOURSTAT. In the former, nights spent are split over the months when they occur, while in the latter they are counted once in the month of departure.

## **Concept 16 – Cost and Burden**

The burden on the respondents is very low since the survey takes only a few minutes to be completed.

Costs associated with the data collection and production of the Tourstat include:

- Investment of tablets
- IT costs for data entry program. The Tourstat program is regularly maintained.
- Twelve full-time survey interviewers, split in two teams, which work daily on a shift basis.
- Uniforms of the twelve survey interviewers.
- Office rent at Malta International Airport (MIA).
- Premises maintenance at MIA.
- Costs incurred for internet, telephone and electricity at MIA.
- A person at the office responsible for the queries of the interviewers and verification of the data.
- Two full-time statisticians that are responsible for the data cleaning/analysis, compilation of news releases, requests and for the final statistical product.

## **Concept 17 – Data revision**

### **Sub-Concept 17.1: Data revision – Policy**

At the NSO, there is currently no internal policy governing revisions that occur for all statistics produced. Nonetheless, a revisions policy is being drafted to safeguard a coordinated revisions system across statistical domains.

This policy will take account of the need and causes for revisions; time and frequency of revisions; data and other statistical products affected by such revisions; and length of periods revised.

### **Sub-Concept 17.2: Data revision – Practice and Data Revision**

Data revision average size – 0%

Each year Tourstat methodology is assessed and any necessary enhancements are put into effect for an improved method. Examples of improvements include new imputation methods for item non-response, new outlier detection techniques and the addition of new questions in the Tourstat questionnaire.

## **Concept 18 – Statistical Processing**

### **Sub-Concept 18.1: Source data**

Source data:

An ongoing frontier survey on inbound and outbound tourism trips and same-day visits.

Survey vehicle:

Stand-alone survey

Population frame:

MIA departing passengers

Quality issues pertaining to the sampling frame (under/over coverage, duplicate records, time lag):

Travellers are only counted once enroute to the departures lounge. To avoid over-coverage, a question asking the foreign passengers if this is their end of visit or if they will return to continue their visit to Malta is asked. A respondent is classified as a returning passenger if: While he/she is in Malta, decided to go for a few days abroad and will be returning to continue his/her visit in Malta. The main aim of this question is to avoid double counting of departing passengers.

All those going through departures at the airport during an interviewing period may be invited to take part in the survey. This description excludes:

- Employees on duty: Including airline/ shipping crew and MIA/Seaport personnel;
- Passengers and crew on domestic flights;
- Re-crosses – i.e. Passengers who have already crossed the counting line on the sample day even if it was before the shift;
- Cancelled flights – If a flight is cancelled and the team leader is able to find out from desk staff that all the passengers have been re-booked airside onto another flight/sailing (i.e. they would not come through the line again), the interviews should be left as eligible. If the passengers will leave the airport/seaport and will have to pass through passport control again, then all passengers on that flight/sailing are ineligible (re-crosses);
- Passengers are eligible if they are departing on a flight which has been delayed overnight;
- Passengers who are refused admission and/or are being escorted through Immigration controls by security or police officers are deemed ineligible.

There exists no time lag between the reference period of the actual population and the sample.

Gross sample size - 88,357 passengers

Net sample size – 54,202 passengers

Reporting unit:

Inbound tourism trips, which comprise of activities of non-resident visitors travelling to Malta (i.e. outside their usual environment) and staying for not more than 12 consecutive months for leisure, business or other (corresponding) purposes.

Outbound tourism trips, which comprises the activities of residents of Malta travelling to, and staying in places outside Malta (i.e. outside their usual environment) and staying for not more than 12 consecutive months for personal, business or other purposes.

Sampling design:

A two-stage sampling design is used to collect air passengers. In the first stage, alternate days and nights are selected. In a fortnight all days and nights are covered. In the second stage, within each shift, a sample of passengers is selected systematically. Every crossing passenger is counted, and respondents are selected using a pre-defined interval of 1:20.

## **Sub-Concept 18.2: Frequency of data collection**

Daily.

### Sub-Concept 18.3: Data Collection

#### Collection procedure:

- The number of interviews to be affected each day is defined by the 1:20 mechanism and depends on the daily flight schedules and the number of departed passengers that were clicked.
- There may be times when a large number of persons cross the line, and this might result in a less accurate count of passing passengers. In such a case counting continues to the best of the interviewer in charge. It is important for the interviewer in charge of the clicker that there is no double counting and that no one is missed.
- When the count reaches the next selection number, the selected passenger who should be approached for interview is selected.
- Interviewers are required to give a brief introduction before conducting each face-to-face interview. Departing passengers are informed that they have been randomly selected to participate in the survey and that information is strictly confidential.
- When the identified passenger refuses to participate in the survey, the next identified passenger (in terms of the 20<sup>th</sup> count) is then approached.
- Method used to gather data from respondents is CAPI.

#### Preparation for Data Analysis:

The interviewers' monthly response rate is closely monitored by the Data Management Unit in order to assess performance. The total number of surveys distributed by the outcome of the questionnaire is checked for each interviewer. A survey is classified as a non-response if it is either clicked, refused or non-contact, whereas responses include complete, partial and minimum outcome questionnaires.

In addition, every month the sampling rate is checked to ensure that every 20<sup>th</sup> passenger was interviewed through each day and night shift respectively. This is carried out by multiplying the total surveys per day by the sampling rate (20) and then subtracting from the total number of passengers that were clicked during that day. If the difference is less than 20 passengers, then this verifies that the sampling rate was held constant throughout.

To make sure that each process is completed in a timely manner, this is the work flow from data collection to dissemination:

1. Monthly updates are done on the lists of airports, airlines and establishments by adding new airports etc. in the current lists.
2. A person is in charge to update the exchange rates daily in order to be used to convert the expenditures in foreign currencies.
3. Data are extracted from the CAPI by the Data Management Unit, which is then sent to the Methodology and Quality unit to import it into SPSS and validate the data.
4. The statistician requests auxiliary data from external sources, 15 days prior to the dissemination of the news release.
5. All releases are published and disseminated at 11:00 a.m. as scheduled in the Advance Release Calendar.

#### Data Analysis:

Process of data analysis:

1. Data validation; process of monitoring the results of data compilation and ensuring the quality of statistical results. Different imputation methods are used in order to deal with item non-response and outlier detection techniques to remove influential points from the dataset.
2. Once the data cleaning process is ready, the Statistician starts working on the weighting procedure. The Statistician must make sure that the preparation of the news release (including tables and commentary) is completed in a timely manner (5 days prior the publication date) so that the Unit Manager and the Director have enough time to check the news release. The news release must be ready and sent both to the Dissemination Unit and to the Tourism Panel, 3 days prior the publication date.

Auxiliary Data requested by NSO:

To achieve the outputs mentioned in sub-concept 1.3, the following auxiliary information needs to be requested by email, 15 days prior the dissemination of Inbound Tourism news release:

- Exchange Rates – Central Bank of Malta (CBM) and the European Central Bank (ECB)
- Flight details - Malta International Airport (MIA)
- Passenger Movements - Malta Tourism Authority (MTA)
- Number of transit passengers - Malta International Airport (MIA)
- Overnight cruise passengers – Transport Malta (TM)
- Sea passenger’s data – Virtu Ferries

All administrative data are received via e-mail in Excel format. Administrative data are compared with the corresponding month of the previous year to check for any inconsistencies.

**Sub-Concept 18.4: Data Validation**

The results of data compilation are monitored to ensure the quality of statistical results by the Methodology and Quality Unit and the domain unit.

The important validations include:

1. Checking date of arrival in Malta with the total number of nights spent.
2. Locality stayed longest is checked with the name of establishment.
3. Purpose of visit is checked with the total number of nights spent.
4. The number of nights spent in Gozo is checked with the categorical variable identifying those who visited Gozo/Comino
5. Package trips, use of travel agency or internet to book are checked with the main type of accommodation (Collective/private accommodation).
6. Checking whether important key variables (Country of residence, end of trip or will return to continue this visit, overnight visitor or same-day visitor, age and gender) are all provided.
7. Checking whether the country of destination is coherent with the trip duration.
8. Checking whether the expenditure amount provided is coherent with the trip duration and type of accommodation. These are checked using multiple linear regression models.
9. Are the answers logical and consistent?
10. Are the time periods possible?
11. Controlling of outliers for length of stay and expenditure categories.
12. Airline used and airport where the respondent is going to from Malta are checked with the flight sheets provided by MIA.
13. Is destination possible?

Special attention is given to completed fields on expenditures. The missing values for expenditures are imputed. Hereby, imputation methods are used, and are based on the core variables destination, number of nights spent, and type of accommodation used. A respondent may also, unintentionally, give the wrong answer.

In addition, every month the sampling rate is checked to ensure that every 20th passenger was interviewed through each day and night shift respectively. This is carried out by multiplying the total surveys per day by the sampling rate (20) and then subtracting from the total number of passengers that were clicked during that day. If the difference is less than 20, this verifies that the sampling rate was held constant throughout. If the difference is equal to 20, this would mean that the interviewer missed one person.

### **Sub-Concept 18.5: Data Compilation**

1. Data collection and entry of the surveys: Is done by the interviewers using CAPI, a web-based software designed to cater for the Tourstat survey. This software was developed by the IT unit within the NSO.
2. Auxiliary data requested by the NSO: Administrative Data (MIA passenger movements by Airport) are used to calibrate the Tourstat survey so that the total passengers will be as per MIA totals.
3. Data extracts: A dataset is extracted from the Tourstat program by the Data Management Unit, and then imported into SPSS software, in order to start the process of data cleaning and analysis.
4. Monthly validations of the inputted data: Are done by the Methodology and Quality unit and later the domain unit: including in-built validations in the data entry program and validations during data analysis to check consistencies between the variables in the SPSS dataset.
5. Data cleaning: Questions which involve Nationality, Country of residence and Airports are coded according to ISO-3166 classification. New variables are derived by recoding the original variables through the use of SPSS software. The new variables are derived to be used in the monthly news releases and tailor-made requests.
6. Imputation of item non-response: Is done using the Hot-deck imputation method for categorical variables; and Multiple Linear Regression for the expenditure categories. Three different outlier detection techniques are used to remove outliers in 'Total nights' and 'Expenditure categories', namely, Standardized Residuals, Cook's distance and Leverages.
7. Weighting procedure: Adjust for unit non-response using post-stratification technique and calibration with MIA passenger movement. A detailed description of the weighting procedure is provided in sup-concept 18.6.
8. Analysis including preliminary checks on main indicators: Including total inbound/outbound tourists. total nights spent by type of accommodation and total expenditure categorized by expenditure categories.
9. Quality checks: including non-response, CVs, coherence with MIA and comparability over time though time series databases.
10. Inbound and Outbound news releases: Are disseminated in terms of aggregated data which includes information on the characteristics of the trip, main types of accommodation, main market shares, nights spent and total expenditure.

#### Treatment of Non-Responses:

There are instances in which the respondent refuses to answer any questions and thus his/her country of origin is not known. It goes without saying that first preference remains obtaining the true country of residence of the respondent, however when this is not possible, our interviewers

are instructed from the unit's end to use "language" as a reasonable indicator. So, in the Tourstat program a "Don't Know" option, divided into three parts is provided:

- "Don't Know Maltese" – used when the respondent declined to take part in the survey is Maltese or when you have reason to believe that s/he is a Maltese national (from his behaviour / appearance etc).
- "Don't Know Foreigner" – used when the respondent is communicating in a foreign language or you have reason to believe that s/he is a foreign national (from his behaviour/ appearance etc).
- "Don't Know" – used when absolutely nothing can be deduced from the respondent's appearance / behaviour / language.

### **Sub-Concept 18.5.1: Imputation**

There are several reasons for the data being unavailable. These include the refusal to provide an answer, the inability to provide an answer, and inadequate quality of the provided answer (e.g. implausible, incomplete, inconsistent with answers to other questions, etc.). It can be caused by either the respondent (e.g. refusal) or the interviewer (e.g. failure to record the answer adequately). In cases where the respondent has either supplied erroneous data or provided high-quality information for most variables but for which data on other variables are missing, erroneous or missing data are estimated by using appropriate imputation techniques. Missing values are imputed in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits or plausible checks. The objective is not reproducing the true micro data value, but to establish consistent and reliable data records that yield aggregate estimates that makes sense.

Different imputation methods are used in order to deal with item non-response:

#### Deductive methods

These methods are rather related to heuristics than to modelling. They try to deduct the most logical answer using the available information for the individual. In general, such procedures will be part of the validation checks and not of the non-response treatment. Example: the respondent did not state the number of nights spent in Gozo but from other questions, we know that he didn't visit Gozo/Comino. In this case, the most obvious value with which to impute the number of nights spent in Gozo/Comino is zero.

#### Imputing by the median:

This method consists of imputing missing values by the median observed in the group of respondents in case of numerical variables. Imputation is done by the median rather than by the mean since it is a robust estimator – it is not affected by outliers. Instead of imputing the overall median, usually, the imputation will be performed by considering some study variables (e.g. destination or purpose of the trip). Within each group, the class median is imputed to cases with missing values. The classes may be different for each variable to be imputed. The big advantage of this method is that is very easy to implement. The main drawback is that it may compress the distribution of the survey variables.

Missing values of the number of nights spent by inbound/outbound tourists are imputed using the median of responding units. The imputation of cases is done within levels of strata using the variables: country of residence, purpose of visit, and main means of accommodation, in order to



improve the quality of the imputation. This method is also used on the variable 'number of nights spent in Gozo/Comino'.

#### Hot deck imputation:

Hot deck imputation: For each respondent with a missing value for a specific variable, this value will be imputed with the corresponding value from the previous respondent in the dataset with a valid response on this variable. This method is applied within classes in order to improve the quality of the imputations. After fixing a starting value for each item and each class, each case is processed sequentially. If the case has a missing item, this is replaced by the imputation value from the relevant class. If the item is not missing, the original value is used for imputation of subsequent missing items.

This method is applied on categorical variables, in our case, Gozo/Comino visit, purpose of Visit, frequency of visit, locality in Malta stayed longest, form of transport used, organisation of the trip, booking of the trip, timeshare accommodation, type of accommodation used, hotel category and kind of accommodation service used/will use.

#### Multiple Linear Regression imputation:

Multiple linear regression models are carried out on all expenditure categories. Linear regression is an approach to model the relationship between a scalar dependent variable and one or more explanatory variables. The method requires that the values of one or more auxiliary variables are known for both the complete cases on which the variable of interest is recorded and for the missing cases. A linear regression model relating the variable of interest to the set of auxiliary variables is set up.

The independent variables which are used in the linear regression models to estimate the missing values in our expenditure categories are the following:

##### 1. Package Expenditure

*Dependent variable: Package Expenditure*

Independent variables: Airline type (Low cost airlines/Other airlines/Sea operator, Type of accommodation used/will use (Hotel/Other collective/Private accommodation) and Nights spent in Malta.

##### 2. Travel Ticket Expenditure

*Dependent variable: Air/Sea Expenditure*

Independent variables: Airline type (Low cost airlines/Other airlines/Sea operator), Country of residence, Destination and Age group.

##### 3. Accommodation Expenditure

*Dependent variable: Accommodation Expenditure*

Independent variables: Purpose of visit (Holiday/Business/Other purpose), Type of accommodation used/will use (Hotel/Other collective/Private accommodation), Nights spent in Malta and Age group.

#### 4. Other Expenditure

*Dependent variable: Other Expenditure*

Independent variables: Country of residence, Nights spent in Malta and Type of accommodation used/will use (Hotel/Other collective/Private accommodation).

Item non-response of the following variables: 'Airport that one embarked from for visiting Malta', 'Airline company', 'Airport that one is going to on this particular trip', 'End/Change airport', 'Airport where the journey will end' is treated by using the flight details sheets (Administrative data) provided on a monthly basis from the Malta International Airport (MIA).

The 'old' variables (containing missing information) and 'imputed' variables (completed data) are both kept in the dataset for all the variables.

#### 5. Outlier detection techniques

Two variables that are most likely to be affected with outliers are the number of nights spent and the expenditure categories. The standardized residuals, which are obtained by fitting a multiple linear regression model, are the most commonly used measures for detecting outliers. Observations with resulting standardized residuals exceeding +2 or -2, indicates that one should investigate these data points. An alternative measure that is used for indicating data points that are particularly worth checking for validity is Cook's distance. A conventional cut-off value that is used for spotting highly influential points is when the resulting Cook's distance is greater than  $4/n$ , where  $n$  is the sample size. The two methods are used to detect outliers for the variables: nights spent and for the different expenditure categories. Another method that is used to detect outliers for the variable number of nights spent is by computing the median of the entire observations and within different strata using the variables which are related with the number of nights spent, that is, purpose of visit and type accommodation used. Using the rule of thumb, that is, if the number of nights spent is greater than the median multiplied by 8, this will indicate that this observation is an outlier. Consequently, these observations will be removed from the entire dataset. Note that for the rule of thumb we are multiplying by a factor of 8 since the average number of nights spent in Malta is 8.

#### 6. Imputation - Rates

Imputation rates for the main variables are approximately as follows:

- Purpose of visit – 2.8%
- Organisation of stay (Package/Non-Package Trip) – 8.6%
- Frequency of visit – 4.0%
- Nights spent - 1.1%
- Package expenditure – 26.0%
- Travel ticket expenditure – 27.8%
- Accommodation expenditure – 27.3%
- Other expenditure – 23.7%

## Sub-Concept 18.6: Adjustment

### Weighting Procedure

- Weights are compiled separately for Maltese and Foreign, and day and night respondents depending on the 1<sup>st</sup> stage of sampling.
- Weight is compiled in 3 stages:

(i) Adjustment for unit non-response:

Maltese and foreigners are weighted separately to adjust for non-response. First, Maltese and foreign respondents from the net sample are distributed by day and night respondents. Similarly, the total number of Maltese and foreign non-respondents are distributed by day and night respondents. The non-respondents with unknown country of residence are classified as Maltese and foreign by applying the same distribution proportion of the non-respondents with known country of residence. Adding the respondents and non-respondents will result in the total gross sample distributed by day and night respondents on the first stage of sampling. To adjust for non-response, the initial weight is then computed by dividing the figures of the gross sample by the respondents of the net sample.

(ii) Sampling rate for the second stage of sampling:

To estimate the number of passengers, the initial weight is multiplied by the inverse of the original sampling rate and by the respondents of the net sample distributed by day and night respondents. The inverse of the original sampling rate is computed separately for each day and night shift, by first dividing the number of days in that particular month by the number of day/night shifts then multiplied by 20, the sampling rate.

(iii) Calibration of weights according to MIA and Valletta Cruise Port departures: Calibration is computed manually in Microsoft Excel: dividing the administrative records by our estimates. Administrative records provided by Malta Tourism Authority (MTA) are used as an auxiliary source for the calibration of the survey aggregates. The exercise is in the main a bottom-up approach including a calibrating weight to re-align the survey results to the true population of total passenger departures (net of transits). Calibration is based on the first country of destination which is not the same as the final country of destination by dividing the true population of total passenger departures by the survey estimates. A correction factor is applied in order to adjust for the festive and non-festive days.

### Sub-Concept 18.6.1: Seasonal Adjustment

Not applicable.

### Concept 19 – Comment

The Tourstat questionnaire may be found on:

[https://nso.gov.mt/en/nso/Sources\\_and\\_Methods/Unit\\_C3/Tourism\\_Statistics/Pages/TOURSTAT\\_Survey.aspx](https://nso.gov.mt/en/nso/Sources_and_Methods/Unit_C3/Tourism_Statistics/Pages/TOURSTAT_Survey.aspx)