TWINNING CONTRACT

BA 15 IPA SR 01 17

Support to the reform of the statistics system in Bosnia and Herzegovina

MISSION REPORT

Activity 2.5.6:
First results from regular survey on tourism statistics

Component 2: Business Statistics
Sub-component 2.5: Tourism Statistics

Mission carried out by
Else Marie Rasmussen, Statistics Denmark
Ivana Brozović, Croatian Bureau of Statistics

28-31 Oct 2019

Version: Final

Expert contact information

Else Marie Rasmussen
Statistics Denmark
Copenhagen, Denmark
Tel: +45 39173029
Email: emr@dst.dk

Ivana Brozović
Croatian Bureau of Statistics
Zagreb, Croatia
Tel: +358 014806272
Email: brozovici@dzs.hr

Table of contents

List of Abbreviations

BHAS	Agency for Statistics of Bosnia and Herzegovina
BiH	Bosnia and Herzegovina
CBBH	Central Bank of Bosnia and Herzegovina
EC	European Commission
EU	European Union
FBiH	Federation of Bosnia and Herzegovina
FIS	Institute for Statistics of Federation of Bosnia and Herzegovina
MS	EU Member State
RSIS	Institute for Statistics of Republika Srpska
RTA	Resident Twinning Adviser
ToR	Terms of Reference

1. General comments

This mission report was prepared within the EU Twinning Project "Support to the reform of the statistics system in Bosnia and Herzegovina". It was the sixth mission to be devoted to 2.5: *Tourism Statistics* within Component 2: *Business Statistics* of the project.

The purposes of the mission were:

- Follow-up from the previous mission
- Prepared by the MS experts
    - Presentation on macro validation
    - Presentation on analysis of aggregate and micro data

- Prepared by the BC experts
    - Clean data from both data collection waves in entities
    - Clean and merge data from the two data collection waves into one
    - Data quality analysis on merged data (response rate, missing values, anomalies)
    - Outlier detection followed by imputation on merged data
    - Grossing-up procedure

- Calculation of first results from regular survey on tourism activity for 2018
- Presentation on quality indicators
- Agreement on output tables (publication forms, transmission formats, etc.)
- Presentation on metadata
- Development of draft methodology
- Preparation of draft quality report

The consultants would like to express their thanks to all officials and individuals met for the kind support and valuable information which they received during the stay in Bosnia-Herzegovina and which highly facilitated the work of the consultants.

This views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, BHAS, FIS, RSIS, CBBH, Statistics Denmark, INSEE, Statistics Finland and Croatian Bureau of Statistics.

2. Assessment and results

The mission began with an introduction of the agenda and a follow-up on the last mission. This was followed by presentations on macro validation prepared by the MS experts.

Macro validation

In general, macro validation consists of comparison over time and comparison with other sources in order to see if the collected data is valid. The MS experts compare the aggregated data with previous years, and this will also be possible for BiH when the survey data for 2019 is ready. The collected data for 2018 can be compared with other sources such as the domestic accommodation statistics as well as accommodation statistics for neighbouring countries. The data for 2018 can also be compared to sources on transport if such are available.

Cleaning of data from both data collection waves in entities

After the presentation on macro validation, some perspectives on the data collected for 2018 were shared. Overall, the data from the second data collection wave covering the period of November-December 2018 is regarded of better quality than the data from the first data collection wave covering the period of January-October 2018.

Data from both data collection waves has been cleaned prior to this mission as planned. There were some challenges regarding different understandings on how to clean data between entities, but the same logical controls as stated in a previous mission were used. Cleaning data consists mainly in

4

recoding wrong country codes or missing codes and similar. Imputation on expenditure and similar is not recommended to be done before the data has been merged into one data set. A larger sample will give a better distribution. It is also recommended to have the outlier detection prior to imputation, but after the merging of the data into one data set.

The cleaned data from the entities has been merged into two data sets covering the whole of BiH – one for the period of January-October 2018 and one for the period of November-December 2018. The main problem with merging the entity data was the use of different data processing systems (IST and Blaise). This caused difficulties with mismatching when the data had to be merged into one data base. Limited resources have also been a challenge. In the second data collection wave most of the difficulties with systemic differences have been eliminated.

Merging of data from the two data collection waves into one

The data from the two data collection waves has not yet been merged into one final data set as scheduled in agenda. The BC experts still need to check if it is feasible to merge the two data collection waves without any more corrections. There are some differences between the two data collection waves:

- The households are not the same in the first and second data collection wave
- The second wave has a participation module
- There are some revisions in questions and modalities.

Some recoding therefore has to be done in order to match the data from the two waves. BHAS will try to merge the two data collection waves during this mission.

Data quality analysis, outlier detection, imputation and grossing-up procedure

The data quality analysis has been done separately for the two data collection waves, but it still needs to be done for the merged data. The same still needs to be done regarding outlier detection followed by imputation on the merged data. Furthermore, the grossing-up procedure also has to finished.

The MS experts made a presentation on examples on outlier detection, imputation and the use of proxy adjusting factors to the weights of trips. It was suggested by the MS experts to consult their sample departments in order to get guidance on how to best adjust weights by using proxy factors. This will be done prior to the next mission.

Data output

Due to the lack of a merged data set, no calculation of first results for 2018 has been done. This has to be done before the next mission.

Discussion on data

The second day of the mission was rounded up by a discussion on the collected data so far. Some logical controls were conducted on the data. There was a question regarding how to manage missing respondents in the participation module. In some cases, only one person from a household consisting of four members for instance answered the questions on participation. The BC experts would like to know how to handle the missing respondents. Are they to be viewed as a unit non-response or an item non-response? The MS experts suggested to look at experiences from similar household surveys. Denmark and Croatia only have individual surveys, so they recommended to consult some of the national statistical agencies where they have a demand side survey with household as the responding unit. BHAS will therefore consult Poland and perhaps Italy, if necessary.

Output tables and metadata

The third day began with a presentation on output tables for national use and for Eurostat made by the MS experts. The rest of the items on the agenda were postponed for the next mission as the merged data is not ready yet. When the merged data is ready, it will be followed by an agreement on the relevant output tables, and the metadata will be discussed in more detail. Subsequently, there were no elaborate presentations on quality indicators and metadata. However, the subject was briefly mentioned by the

MS experts in their presentations on output tables. The MS experts recommended to adhere to the metadata templates, quality reports and transmission formats already provided by Eurostat.

A draft methodology will be developed prior to the next mission and discussed in the next mission. A draft quality report based on the QPI standards provided by Eurostat will also be drafted by the BC experts prior to the next mission.

3. Conclusions and recommendations

During this mission there was a presentation on macro validation by the MS experts followed by a discussion on how to macro validate data and also the use of outlier detection and imputation. The first results from the survey have not yet been produced, as the data from the two data collection waves still has to be merged into one. The MS experts recommend that the merging of the data is prioritized above all the other items on the agenda. It is therefore important to have a merged data set and some results prior to the next mission.

The MS experts made a presentation on output tables and also briefly talked about metadata and quality reports. It is in general recommended to use the Eurostat templates for output tables as well as metadata and quality reports as guidelines. Due to the delay regarding a merged data set, no methodology and quality report was drafted. This has to be done prior to the next mission.

The MS experts recommend to primarily focus on producing a basic, preliminary output table on overnight trips broken down by purpose and destination for the reference year 2018. Furthermore, the MS experts recommend using the outlier detection formula provided by Eurostat for the initial outlier detection and only utilize it on the expenditure in order to save time and resources. The missing values can afterwards be imputed by a basic mean imputation method. When the preliminary results have been published in December, a more detailed outlier detection and imputation method can be used to produce further output tables on trips, participation and same-day visits in 2020.

Action	Deadline	Responsible person
Consult the Sample Departments on the use of adjusting factors to the weights	End of November 2019	MS experts
Consult Polish colleague on how to deal with missing answers from several household members.	End of November 2019	BHAS
Merge data from the two data collection waves into one data set	End of November 2019	BHAS, RSIS and FIS
Outlier detection followed by imputation on merged data	December 2019	BHAS, RSIS and FIS
Grossing-up procedure	December 2019	BHAS, RSIS and FIS
Calculation of first results for 2018	December 2019	BHAS, RSIS and FIS
Further data analysis and output tables	February 2020	BHAS, RSIS and FIS
Data quality analysis on merged data	February 2020	BHAS, RSIS and FIS
Development of draft methodology	February 2020	BHAS, RSIS and FIS
Preparation of draft quality report	February 2020	BHAS, RSIS and FIS
Presentation on quality indicators	February 2020	BHAS, RSIS and FIS
Presentation on metadata	February 2020	BHAS, RSIS and FIS

Annex 1. Terms of Reference

<div style="border: 1px solid black;">

Terms of Reference

EU Twinning Project BA 15 IPA ST 01 17

Component 2: Business Statistics
Sub-component 2.5: Tourism Statistics
28 - 31 Oct 2019
Hosting institution: BHAS, Zelenih beretki 26, Sarajevo

</div>

Activity 2.5.6: First results from regular survey on tourism statistics

1. Mandatory result and benchmarks for the component

Mandatory result:

- New indicators on demand-side tourism statistics, in accordance with EU Regulation 692/2011 (Annex II – National tourism) produced and made available to users by 8th project quarter

Benchmarks:

- Plan for development of demand-side tourism statistics produced by 2nd project quarter
- Questionnaire for a regular survey prepared by 2nd project quarter
- Criteria for an IT application defined by 5th project quarter
- First results of survey analyzed by 6th project quarter
- Indicators on demand-side tourism statistics compiled by 7th project quarter
- Indicators on demand-side tourism statistics made available to users by 8th project quarter
- Methodological document on demand-side tourism statistics developed by 8th project quarter
- Quality report for tourism statistics developed by 8th project quarter

2. Purpose of the activity

- Follow up from the previous mission
- Prepared by the MS experts
  - Presentation on macro validation
  - Presentation on analysis of aggregate and micro data
- Prepared by the BC experts
  - Clean data from both data collection waves in entities
  - Clean and merge data from the two data collection waves into one
  - Data quality analysis on merged data (response rate, missing values, anomalies)
  - Outlier detection followed by imputation on merged data

- o Grossing-up procedure
- Calculation of first results from regular survey on tourism activity for 2018
- Presentation on quality indicators
- Agreement on output tables (publication forms, transmission formats, etc.)
- Presentation on metadata
- Development of draft methodology
- Preparation of draft quality report

3. Expected output of the activity
- First results from regular survey produced
- Output tables defined (publication forms, transmission formats, etc.)
- Metadata presented
- Input to a draft methodology made
- Input to draft quality report made
- Input provided to the ToR of next activity

4. Participants
Agency for Statistics of BiH (BHAS)
Alen Bajramović
Azra Bander Demirović

Institute for Statistics of Federation of BiH (FIS)
Dženana Vreto
Edina Mehidić
Nikola Radović

Institute for Statistics of Republika Srpska (RSIS)
Jelena Glamočika
Jelena Kljajić
Vedrana Dejanović
Slađana Nikić

MS Experts
Else Marie Rasmussen, Statistics Denmark
Ivana Brozović, Croatian Bureau of Statistics

Twinning Project Administration
Katja Møller Hjelvang, RTA
Đemka Šahinpašić, RTA Assistant
Svjetlana Pavičić, Interpreter

Annex 2. Most common errors

Most common errors	Suggestion from MS experts
• YES to a trip, but no expenditure given. • YES to rented accommodation, but no accommodation expenditure given. • YES to transport, but no transport expenditure given. • YES to package trip, but no package trip expenditure.	The simple solution is to mark the expenditure as imputed and use the mean imputation method: Using averages grouped by destination (domestic/foreign), purpose (private/business), duration or grouped by background variables such as age, gender, education. The trip characteristics can be combined with the background variables.
The total expenditure is given, but expenditure items are missing.	• If the total expenditure is known, but some expenditure items are unknown or missing, then you can estimate the items by applying the average share by similar respondents or/and similar trips. • If you miss just the expenditure for one item, impute for the difference between the total cost and the sum of the expenditure for items • If missing data on expenditure for more items - the difference between the total expenditure and the sum of the reported expenditure is calculated, and this difference is allocated to items that are proportional to the number of respondents who reported all costs
Missing data on resident or non-resident.	If they filled out the questionnaire, they should be marked as resident.
Trips outside of reference period.	If a trip has ended outside the reference period, it should be filtered out.
YES to private trips, but no destination given.	If no destination is given, you can choose the most probable answer by defining some criteria such as "If Transport=By plane" AND duration > 6 nights, then destination=MOST visited foreign destination with more than > 6 nights and transport by air. In other words, the most frequent answer given grouped by relevant variables = Mode imputation.
Expenditure in Q24 and Q25 differs.	If the expenditure differs, choose the expenditure estimate in Q25 and estimate the expenditure items by applying the shares for the expenditure items for similar trips or respondents.
Number of trips is given, but not all of the trips are listed with characteristics.	It is possible to use the listed trip characteristics as a proxy for the missing trip characteristics. This is however not recommended in a large scale.
YES to same-day visit, but no purpose of the same-day visit listed.	Use mode imputation or logical rule such as: If trip is done alone, it could be business, if done with household members, then it is private.
YES to trip, but no overnights stays filled in.	Mode imputation grouped by destination and purpose.

Annex 3. Validation controls

- **Socio-demographic questions**
  - o Q1 on residency in BiH last 12 months: If not 1 or 2, then error.
  - o Q2 on name: String variable.
  - o Q3 on gender: If not 1 or 2, then error.
  - o Q4 on relation: If not 1-5, then error.
  - o Q5 on year of birth: If not 1900 or higher until current year, then error. Should be length 4.
  - o Q6 on education: Values 1-6, "Refusal" or "Not applicable".
  - o Q7 on employment status: Values 1-4, "Refusal" or "Not applicable".
- **Number of trips**

  ID-number of person speaking on behalf of the household. Maximum is the number of household members, and children are to be excluded.
  - o Q1: Values 1,2. If 2, then go to Q6.
  - o Q2: Values 0-99. If 0, then go Q4.
  - o Q3: Values 0-99. Q3 has to be =<Q2.
  - o Q4: Values 0-99. If 0 then go to 6.
  - o Q5: Values 0-99. Q5 has to be =<Q4.
  - o Q6: Values 1,2.
  - o Q7: Values 0-99. Q6 has to be 1. If it is 0 then go to Q9.
  - o Q8: Values 0-9. Q8 has to be =<Q7. If higher than 9, then check plausibility
  - o Q9: Values 0-99.
  - o Q10: Values 0-9. Q10 has to be =<Q9. If higher than 9, then check plausibility
  - o Q11: Values 1-8 or "Refusal" or "Don´t know". If Q1=2 or Q2=0, then Q11 should be answered. Should be moved to the participation module.
- **Characteristics on trips**
  - o Q1: String variable. Dropdown list in database.
  - o Q2: If Q1="BiH", then answer this question. Answer with city/municipal-code.
  - o Q3: Format MM YY. Link to reference period (T and T-1)
  - o Q4: Number of nights: Values 1-366.
  - o Q5: Choose maximum on the basis of what data shows. Q5 should be < than Q4
  - o Q6: Only answer. Values 1-14 and "Refusal" and "Don´t know". Link the type of trip under number of trips with purpose of trips if possible. Example, if no business trips, then the purpose should be private. If Q6=private, then answer Q7. Else go to Q8.
  - o Q7: Multiple answers possible for 1-6 or "Don´t know" and "Refusal".
  - o Q8: Transport. Values 1-6 and "Don´t know".
  - o Q9: Booking transport by travel agent: Values 1,2 and "Don´t know".
  - o Q10: Booking transport online: Values 1,2 and "Don´t know".
  - o Q11: Type of accommodation: Values 1-8, "Refusal" and "Don´t know". If non-rented in (6,7,8), then go to Q14.
  - o Q12: Booking of accommodation by travel agent: Values 1,2, and "Don´t know". If non-rented, then Q12=2.
  - o Q13: Booking of accommodation online: Values 1,2, and "Don´t know". If non-rented, then Q13=2.
  - o Q14: Other mode of booking. Only to be answered if Q9=2 and/or Q12=2.
  - o Q15: Expenditure participants should correspond with the travel party size. Change in IG, either individual or whole travel party.
  - o Q16: If Q9 =1 or Q12 =1, then answer this question. Values 1,2, and "Don´t know". If value=1, then go Q17, if value=2 or don´t know, go to Q19.
  - o Q17: Total cost of package trip. Only to be answered, if Q16=1. Values 1-99999.
  - o Q18: At least two answers should be chosen. Multiple answers.
  - o Q19: Transport expenditure: Values 0-9999. Don´t know and refusal. If Q16=2, then go to Q20, else go to Q19a.
  - o Q19a: Additional transport expenditure apart from package trip expenditure on transport. Q18=1.

- o Q20: Accommodation expenditure: Values 0-9999. Don´t know and refusal. If Q16=2, then go to Q21, else go to Q20a.
- o Q20a: Additional accommodation expenditure apart from package trip expenditure on accommodation. Q18=1.
- o Q21: Expenditure on food and drinks: Values 0-9999. Don´t know and refusal. If Q16=2, then go to Q22, else go to Q21a.
- o Q21a: Additional expenditure on food and drinks apart from package trip expenditure on food and drinks. Q18=1.
- o Q22: Other expenditure: Values 0-9999. Don´t know and refusal. If Q16=2, then go to Q22, else go to Q22a.
- o Q22a: Additional other expenditure apart from package trip expenditure on other. Q18=1.
- o Q23: Expenditure on durable and valuable goods. Q23 has to be <= to Q22 or Q22a. Minimum of 300 KM.
- o Q24: Checking total: Total=Transport + Accommodation + Food and drinks + Other. If package trip, then total=total cost of package + additional transport + additional accommodation + additional food and drinks + additional other. The value reported is to be the individual expenditure, so if the respondent answers on behalf of the whole travel party, then the expenditure should be divided by the number of participants.
- o Q25: If they did not answer the separate expenditure, then please estimate a total expenditure. If Q24=0, then answer Q25. The estimated total expenditure should be broken down by categories by using the proportion of the other responses grouped by similar respondents.
- **Same-day visits characteristics**
    - o Copy from trips
- **Income question**
    - o Q99: Values 1-8, "Refusal" and "Don´t know"
- **Participation module**
- **Correlation between questions (deductive method)**
    - o Calculate data minimum, maximum, median, mean value. For example put median number of trips as threshold. If higher, then put a warning.
    - o If number of trips is <=5, then number of columns under characteristics of trips should be = to that.
    - o If the number of trips is higher than 5, then the weight shall be multiplied with Number of trips total/Number of trips in detail. Group by relevant variables. Proportional principle.
    - o If they have at least 1 personal trip, then purpose should be 1-12.
    - o If they have at least 1 business trip, then purpose should be 13,14
    - o Distribution of domestic and foreign destinations shall correspond on answers on trips home and abroad under number of trips. General rule in the columns.
    - o Transport and expenditure
    - o Transport and destination. If outside Europe, then by air if missing. What is the most frequent value when travelling in BiH. Use this as imputation measure if missing.
    - o If package trip then accommodation has to be rented.
    - o If accommodation is rented, then accommodation expenditure should be higher than 0, if they have answered the separate category.
    - o If accommodation is non-rented then accommodation expenditure should be 0.
    - o If package trip then yes to use of travel agent.
    - o If package trip, then package trip expenditure > 0.
    - o If transport mode 1-5 and no to package trip, then transport should be higher than 0.
    - o Correlation between annual number of trips and annual number of participants. Number of participants should always be =< than number of trips.
    - o Outlier detection for accommodation expenditure is calculated by looking at accommodation expenditure/ number of nights spent.
    - o Correlation between household members and participants in the trip.

- o   Similar controls for same-day visits
  - o   Match answers in number of trips last quarter with the participation module.
- **Outlier detection: Calculate possible outliers and look at them manually.**
- **Imputation or keep.**