

TWINNING CONTRACT

JO/13/ENP/ST/23

Strengthening the capabilities of the Department of Statistics in Jordan



MISSION REPORT

on

Activity: 2.8 Imputation and imputation using R

Mission carried out by
Mr. Ugo Guarnera, ISTAT
Mr. Kai Lorentz, Destatis

7–11 December 2014

Version: Final



Expert contact information

Dr. Ugo Guarnera
Head of the unit
“Treatment of non-sampling error”
ISTAT
Via Cesare Balbo 16
00184 – Roma
Italy
Tel. +39 06 46731
Email: guarnera@istat.it

Dr. Kai Lorentz
Mathematical Statistical Methods
Destatis
Gustav-Stresemann-Ring 11
65189 Wiesbaden
Germany
Tel: +49-(0)611-75 25 89
Email: kai.lorentz@destatis.de

Table of contents

1. General comments	4
2. Assessment and results	4
3. Conclusions and recommendations	6
Annex 1. Terms of Reference.....	7
Annex 2. Persons met.....	11
Annex 3. Training material.....	12

List of Abbreviations

DoS	Department of Statistics of Jordan
ToR	Terms of Reference

1. General comments

This mission report was prepared within the Twinning Project "Strengthening the capabilities of the Department of Statistics in Jordan". It was the 6th mission within Component 2 "Sampling techniques" of the project.

The purposes of the mission were:

- To introduce to DoS staff members the concepts of imputation methodology.
- To explain and train DoS staff members on the usage of the "R"-software, especially in the context of imputation.
- To provide consultation in imputation problems for the HIES (household income and expenditure survey), the Industry Survey and Agricultural Surveys of the Kingdom of Jordan.
- To tackle also specific questions from DoS related to sampling methodology.

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 Jordan and which highly facilitated the work of the consultant.

This views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, DoS, ISTAT or Destatis.

2. Assessment and results

Starting point is the list of activities mentioned in the ToR (see Annex 1).

A five days training – combined with consultations – took place at DoS of Jordan from 7th to 11th of December 2014.

The topics have been the following:

- A general introduction into imputation methodology.
- Demonstration of examples on imputation methods using the "R"-software and related packages.
- Dealing with data files; importing and exporting files from Excel and SPSS into "R"-software.
- Dealing with missing values for the Jordanian HIES (Household Income and Expenditure Survey), general recommendations and treatment with the "R"-software.
- Dealing of missing values within the Jordanian Industry Survey, general recommendations and treatment with the "R"-software.
- Drawing of a sample for the Jordanian Household Survey (region of Aqaba as an example) with the help of the "R"-software.
- General issues on imputation for household surveys have been discussed with the "Household Survey Division" of DoS.
- Discussions on a possible sampling plan for capital formation survey within the agricultural sector, to be developed in the near future.

Relevant course material (presentations, program scripts and example data) can be found in Annex 3 of this report and had been handed over within the mission to DoS staff members.

The training and consultations took place in a very friendly atmosphere. The training centre had been perfectly prepared with the installation of the required software.

The participants were very active in asking questions on general statistical methodology and on technical issues of the presented software. It was recognized by the participants that the R-software is of great practical relevance for some of the staff members of DoS, as an add-on to the existing IT environment.

The participants were introduced into the general concepts on imputation. Treatment of non-response is of practical relevance for DoS, both unit and item non-response.

The effects of different imputation methods for different kinds of missingness of data values has been demonstrated with the help of examples, demonstrating at the same time the usage of “R”-software.

Missing values within the Household Income and Expenditure survey (HIES) has been treated with the help of the “R”-software. A relevant data file had been sent to the consultants in advance. However, it turned out in the discussions, that almost all cases of non-response are unit non-response, so that re-weighting would be the appropriate method of treatment. Nevertheless, a solution using a hotdeck imputation method was also discussed and worked out with the “R”-software during the mission.

Similar to the HIES, missing values within the Jordanian Industry Survey has been treated within the mission. Also in this case, a data file had been sent in advance to the consultants. Scripts for data treatment and imputation within “R” have been elaborated during the mission.

As an answer to a request from participants of the training, the drawing of a sample with the help of the “R”-software has been demonstrated. As an example data frame served census data from the region of Aqaba.

General issues on imputation for household surveys have been discussed with representatives from DoS. Of special interest has been the question on missing quarterly values for households for the Expenditure Survey.

The Division of Agricultural Statistics presented to the consultants their need to have a sampling plan developed in the near future for the Survey on “Capital formation in the agricultural sector”, to be used by National Accounts of Jordan.

It is required to produce monthly data on this topic. Currently, monthly data is available only for crop production but not for livestock production.

So a redesign of the survey methodology is required, producing reliable results (95% coverage interval is intended) for the major asset indicators/variables.

As a sampling frame data from the 2007 agricultural census should be used.

3. Conclusions and recommendations

The usage of the R-software as an additional tool within the IT environment of DoS has been general accepted. But, the usage of this software is only beneficiary for specialized staff members.

This group of staff members has to get more acquainted with the usage of this software. It is recommended to apply this software more often, especially in the context of sampling plan development, drawing of a sample, point estimation and variance estimation (CV calculation) as well as imputation.

Concerning imputation, the following very general recommendations are given:

- DoS staff members have to analyse carefully, whether blanks in the data actually represent the value “0” or real missings. If the value “0” is meant by the blank, “0” should be at the corresponding places instead of a blank.
- One should distinguish the cases of unit non-response (whole unit did not provide answers) from the cases of item non-response (only the values of some variables are missing). The first case should in general be treated with re-weighting, the second with imputation.
- Reasons of non-response have to be analysed carefully. If a unit for example does not exist any more (e.g. a company which stopped their economic activity), nothing has to be imputed. In this case, imputation would lead to an over-estimation of the considered variables.
- For economic surveys, imputation cells according to size class and/or economic activity should be considered.

Concerning the planned survey on capital formation within the agricultural sector, it is recommended to develop a sampling plan in a future mission, if time and resources are available.

Action	Deadline	Responsible person
Development of a sampling plan for the capital formation survey within the agricultural sector.		Consultant of a future expert mission (if time and resources are available).
Getting more acquainted at DoS with the usage of the R-software.		DoS staff members.

Annex 1. Terms of Reference and Programme

Terms of Reference

EU Twinning Project JO/13/ENP/ST/23

7-11 December 2014

Component 2: Sampling techniques

Activity 2.8: Imputation and imputation using R

0. Mandatory results and benchmarks for the component

- Improve the capacity of DoS staff to understand and apply modern sampling techniques (Apr 2015)
- Assessment report on current situation (Jan 2013)
- Provide inputs to the design of surveys (Aug 2014)
- Conduct a training course in seasonal adjustment (Oct 2014)
- Give recommendations on how to deal with weights, imputation, non-response and sampling errors (Apr 2015)

1. Purpose of the activity

To discuss the methodology for imputation
To discuss imputation using the R software
To introduce the R packages for imputation
To apply the methods on relevant DoS data

2. Expected output of the activity

DoS staff introduced to and trained in the use of the software R
DoS staff trained in the theory behind imputation
DoS staff introduced to the R packages for imputation
Methods applied on relevant DoS data
Transfer of the Italian and German, and in general the European Union, experience in the use of R-software for imputation

3. Participants

DoS

Mrs Ghaida Khasawneh, Head of Sampling Division (*Component Leader*)

Staff from the Directorate of statistics studies

Staff from the Directorate of Economic surveys

Staff from the Directorate of Agricultural statistics

Staff from the Directorate of Household Surveys

MS experts

Mr. Kai Lorentz, Head of section “Mathematical-statistical methods”, DeStatis

Mr. Ugo Guarnera, Head of the unit “Treatment of non-sampling error”, ISTAT

Programme for the mission

Time	Place	Event	Purpose / detail
Sunday, morning	08.30 – 09.30	Hotel /DoS	Meeting with RTA To discuss the programme of the week
Sunday, morning	09.30 – 12.00	DoS	Meeting with BC Component Leader and BC Experts Initial discussion on Imputation and the R software (imputation related) Repetition and further explanations on the R and the R Studio software
	12.00 – 01.00		Break / Preparations / Report writing Break / Preparations / Report writing
Sunday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts Continued.
	03.30 – 04.00		Preparations / Report writing Preparations / Report writing
Monday, morning	08.30 – 09.30	DoS	Preparations / Report writing Preparations / Report writing
	09.30 – 12.00		Meeting with BC Component Leader and BC Experts Discussion of the concepts of imputation and the terminology. Based on DoS' data and requirements.
	12.00 – 01.00		Break / Preparations / Report writing Break / Preparations / Report writing
Monday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts Continued.
	03.30 – 04.00		Preparations / Report writing Preparations / Report writing
Tuesday, morning	08.30 – 09.30	DoS	Preparations / Report writing Preparations / Report writing
	09.00 – 12.00		Meeting with BC Component Leader and BC Experts Presentation of the R packages for imputation.
	12.00 – 01.00		Break / Preparations / Report writing Break / Preparations / Report writing
Tuesday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts Continued
	03.30 – 04.00		Preparations / Report writing Preparations / Report writing

Wednesday, morning	08.30 – 09.30	DoS	Preparations / Report writing	Preparations / Report writing
	09.30 – 12.00		Meeting with BC Component Leader and BC Experts	Based on DoS datasets, development of a script for imputation. The script should cover: -Reading the data into R -Conducting the imputation -Exporting the data -Analysing the results
	12.00 – 01.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Wednesday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Continued
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Thursday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 11.30		Meeting with BC Component Leader and BC Experts	Summing up the outcomes. Discussions on open problems.
			Ad-hoc meetings	Final clarifications with BC Experts, preparation of report and presentation for BC Project Leader
Thursday, morning	11.30 – 12.30	DoS	Meeting with BC Component Leader	Presentation for BC Project Leader
Thursday, noon	12.30 – 01.00	DoS	Debriefing with BC Project Leader	Conclusions and decisions and their consequences for the next activity and the implied work programme for BC Experts

Annex 2. Persons met

DoS:

Ghaida Khasawneh	Methodologies and Statistical Techniques Directorate
Wafaa Amer	Methodologies And Statistical Techniques Directorate
Abdullah Mostafa	Methodologies And Statistical Techniques Directorate
Mohammad Aljawarneh	Directorate Of Agricultural Statistics
Suzan Al-Owaidy	Directorate Of Agricultural Statistics
Kawther AlAbbadi	Directorate Of Agricultural Statistics
Mohammad Tawalbeh	Directorate Of Agricultural Statistics
Saed Shawawreh	Directorate Of Agricultural Statistics
Zaid Abu Alghanam	Directorate of Household Surveys
Shadi Al Gubilat	Directorate of Household Surveys
Noor Alenaimat	Directorate of Household Surveys
Zaidoon Assad	Directorate Of Economic Surveys
Asharaf AlHajaj	Directorate Of Economic Surveys

RTA Team:

Mr Thomas Olsen, resident twinning adviser (RTA)
Ms Christine Salman, RTA assistant
Interpreter

Annex 3. Training material

The material used during the training can be found in the enclosed ZIP-file.

- A. Presentation on imputation methodology.
- B. R-script containing examples for treatment of different kinds of missingness in survey data together with the example data set.
- C. R-script for treatment of HIES survey data together with the HIES data set.
- D. R-script for treatment of Industry survey data together with the Industry survey data set.
- E. R-script for sample selection together with example data set (Aqaba region).