TWINNING CONTRACT

JO/13/ENP/ST/23

Strengthening the capabilities of the Department of Statistics in Jordan

MISSION REPORT

on

Activity: 2.5 Weighting, imputation, non-response and sampling errors

Mission carried out by
Mr Peter Tibert Stoltze, Statistics Denmark
Mr Kai Lorentz, Destatis

15–19 June 2014

Version: Final

STATISTICS DENMARK

Istat
Istituto Nazionale di Statistica

DOS 1949

Expert contact information

Peter Tibert Stoltze
Deputy Head of Division
Research and Methods
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3827
Email: PSL@dst.dk

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

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 4th mission within Component 2 “Sampling techniques” of the project.

The purposes of the mission were:

- To train DoS staff members in the usage of the R software with emphasis to applications in the area of sampling techniques.
- To provide consultation in weighting problems for the HIES (household income and expenditure survey) of the Kingdom of Jordan.

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, Statistics Denmark or Destatis.

2. Assessment and results

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

A 5 days training took place at the training centre of DoS on the R-software with focus on applications in the area of sampling.

The content of the training has been the following:

- A general introduction to the basics of the R software and the graphical user interface RStudio.
- Explanation of the basic functionality of R on the example of mean-value imputation for missing values (item non-response).
- Usage of functions for sample size calculation from the package “samplingbook”.
- Simulation of drawing samples and estimating, deriving the distribution of sample estimates.
- Usage of function for drawing samples from the package “sampling”.
- Dealing with data files; importing and exporting files from Excel and SPSS into R-software.
- Usage of R-code from Destatis to allocate sample size to strata.
- Usage of the package “survey” for the analysis of survey data; as an example the problem of reweighting for unit non-response for the HIES has been treated.

Relevant course material can be found in Annex 3 of this report.

The training 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.

Of special interest for the audience was the topic of imputation, where only a small example – not really recommended for direct usage at DoS – had been discussed. This topic merits a more intensive dedication within another mission to follow, where some methodological background and practical implementation in R-code should be treated.

A second task of this mission has been the consultation on the subject of reweighting on the household income and expenditure survey (HIES) as treatment of unit non-response.

The HIES is designed as a two-stage household sample (even three-stage if one is looking at the level of individual persons). The design is self-weighted; the first stage selection probabilities are proportional to size (the number of households within the primary sampling unit) and the sample sizes at second stage are all equal to 8. From the selected households, quite often only a subset of households responded, so question of reweighting arises.

The problem was analysed on survey data from the region of Aqaba. It was suggested to apply a calibration method for the adjustment for non-response. The calculations were carried out using the "survey" package and demonstrated during the training. The suggested solution assumes that suitable auxiliary information is available. In the example we use results from a population census and make the estimates from the household survey consistent with these. This is a quite advanced method that should be used with some cautiousness because it can alter the point estimates dramatically. The R-code can also be found in Annex 3.

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. Given the rich sources of information about R on the internet self-study is a good possibility. However, language might be a barrier, since most of this information is in English.

It was emphasized by the consultants, that statistical methodology consists of both statistical understanding and programming proficiency. R is a powerful tool, but one what should be very cautious about the statistical meaning of the operations performed.

It was emphasized by the Jordanian counterparts that gaining knowledge on imputation is of great interest at DoS, in the areas of economic surveys and also household surveys.

So a forthcoming mission on this topic within the component of sampling is recommended.

Action	**Deadline**	**Responsible person**
Definition of a forthcoming mission dedicated to imputation issues.		
Providing the experts with example data files	1 - 2 months before the mission	
Getting more acquainted at DoS with the usage of the R-software		

Annex 1. Terms of Reference

Terms of Reference

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

15-19 June 2014

Component 2: Sampling techniques

Activity 2.5: Weighting, imputation, non-response and sampling errors

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 weighting, imputation, non-response and sampling errors
- To apply the methods on relevant DoS data
- To introduce the software R as an additional tool to be applied by DoS

2. Expected output of the activity

- DoS staff trained in the theory behind weighting, imputation, non-response and sampling error estimation
- Methods applied on relevant DoS data
- DoS staff introduced to and trained in the use of the software R
- Transfer of the Italian and United Kingdom, and in general the European Union, experience in the use of R, weighting, imputation, non-response and sampling errors

3. Participants

DoS

Ghaida Khasawneh	Methodologies and Statistical Techniques Directorate
Ahmad Mowafi	Methodologies and Statistical Techniques Directorate
Abd Alnaser Aljarere	Methodologies and Statistical Techniques Directorate
Wafaa Amer	Methodologies and Statistical Techniques Directorate
Khaled Hrout	Methodologies and Statistical Techniques Directorate
Lina Hasan	Directorate of Economic Surveys
Ensaf Al Nameer	Directorate of Economic Surveys
Faten AlRamini	Directorate of Economic Surveys
Mohammad AlJawarneh	Directorate of Agricultural Statistics
Abd Alnasir obidat	Directorate of Agricultural Statistics
Kawther AlAbbadi	Directorate of Agricultural Statistics
Fatmeh Awamreh	Directorate of Household Surveys
Sana AlMomani	Directorate of Household Surveys
Basman Al-Ajlouni	Directorate of Household Surveys

MS experts

Mr. Peter Stoltze, Statistics Denmark
Mr. Kai Lorentz, Destatis

Programme for the mission

Time		Place	Event	Purpose / detail
Sunday, morning	10.00 – 10.30	Hotel /DoS	Meeting with RTA	To discuss the programme of the week
Sunday, morning	10.30 – 01.00	DoS	Meeting with BC Component Leader and BC Experts	Initial discussion on tasks to calculate with the R software (sampling related) General introduction to R software (part 1)
	01.00 – 02.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Sunday, afternoon	02.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Information and follow-up on activities since activity 2.2 Discussions and presentation of the practical challenges DoS staff would like to discuss with the experts.
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Monday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 01.00		Meeting with BC Component Leader and BC Experts	General introduction to R software (part 2) Introduction and demonstration of package "Stratification" to calculate optimal boundaries on size classes
	01.00 – 02.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Monday, afternoon	01.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Practical work on topics suggested by DoS
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Tuesday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 01.00		Meeting with BC Component Leader and BC Experts	Demonstration of R code "Optalloc" (from DeStatis) for sample size allocation to strata (sampling plan development)
	01.00 – 02.00		Break / Preparations / Report writing	Break / Preparations / Report writing

Tuesday, afternoon	02.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Practical work on topics suggested by DoS
	03.30 – 04.00		Preparations / Report writing	Preparations / Report writing
Wednesday, morning	08.30 – 09.00	DoS	Preparations / Report writing	Preparations / Report writing
	09.00 – 01.00		Meeting with BC Component Leader and BC Experts	Demonstration of package "Survey" for calculating standard errors
	01.00 – 02.00		Break / Preparations / Report writing	Break / Preparations / Report writing
Wednesday, afternoon	02.00 – 03.30	DoS	Meeting with BC Component Leader and BC Experts	Practical work on topics suggested by DoS
	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
Ahmad Mowafi	Methodologies and Statistical Techniques Directorate
Abd Alnaser Aljarere	Methodologies and Statistical Techniques Directorate
Wafaa Amer	Methodologies and Statistical Techniques Directorate
Khaled Hrout	Methodologies and Statistical Techniques Directorate
Lina Hasan	Directorate of Economic Surveys
Ensaf Al Nameer	Directorate of Economic Surveys
Faten AlRamini	Directorate of Economic Surveys
Mohammad AlJawarneh	Directorate of Agricultural Statistics
Abd Alnasir obidat	Directorate of Agricultural Statistics
Kawther AlAbbadi	Directorate of Agricultural Statistics
Fatmeh Awamreh	Directorate of Household Surveys
Sana AlMomani	Directorate of Household Surveys
Basman Al-Ajlouni	Directorate of Household Surveys

RTA Team:

Mr Thomas Olsen, resident twinning adviser (RTA)
Ms Amal Aliah, RTA assistant
Interpreter

Annex 3 Training material

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

A. "A (very) short introduction to R" by Paul Torfs & Claudia Brauer, Wageningen University, The Netherlands.

B. Description of the package "samplingbook".

C. Description of the package "survey".

D. Description and examples with data for the package "OptAlloc".

E. Demonstration of graphics export.

F. Demonstration of sampling proportions (binary variable of interest).

G. Demonstration of sampling continuous variable.

H. Demonstration of how to program a function (case: simple mean imputation)

I. Script for calibration of the HIES with respect to distribution of age groups.

J. Data for the HIES script.

The scripts E-I were prepared by the consultants during the mission. Note that scripts F-H includes suggestions to further work.