

TWINNING CONTRACT
Support to Statistics

Kosovo



MISSION REPORT

on

Methodology Training in Business Statistics

Component No. 3.7.3

Mission carried out by
Peter Tibert Stoltze and Regin Reinert,
Statistics Denmark
9-12 February 2016

IPA 2012

Peter Tibert Stoltze
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3827
Email: psl@dst.dk

Regin Reinert
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3609
Email: reg@dst.dk

Table of contents

1. General comments.....	4
2. Assessment and results.....	4
3. Conclusions and recommendations.....	5
Annex 1. Terms of Reference	6
Annex 2. Persons met at KAS.....	10

List of Abbreviations

DST	Statistics Denmark
KAS	Kosovo Agency for Statistics
NACE	Statistical Classification of Economic Activities
NSI	National Statistical Institute
RTS	Retail Trade Statistics
SBR	Statistical Business Register
SBS	Structural Business Statistics
STS	Short Term Statistics
ToR	Terms of Reference

1. General comments

This mission report was prepared within the Twinning Project „Support to Statistics”. It was the third mission to be devoted to Methodology Training in Business Statistics within Component 3 of the project. The mission was aimed at extending the knowledge of survey sampling methodology and the R software for statistical programming, both for survey sampling estimation and for connection to the business register database at KAS.

The concrete objectives of the mission were:

- Draw a new sample for SBS
- Discuss and train estimation in the presence of non-response using SBS as an example
- Discussing how to use “R” in connection with a database system

The consultants would like to express their sincere thanks to all officials and individuals met for the kind support and valuable information which we received during the stay in Kosovo, and which highly facilitated our work.

The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, KAS or Statistics Denmark.

2. Assessment and results

The Terms of Reference (ToR) stated three activities to be conducted during the mission:

1. Drawing a new sample for SBS
2. Training on selected items of methodology
3. Discussing how to use “R” in connection with a database system

Ad. 1. Due to different circumstances data to create a sampling frame for the new SBS sample were not available yet. However, the procedures used for creating the RTS sample during mission 7.3.2 are quite generic and can be adopted to create a sample for the SBS as well. The structure of the SBS sample was discussed, especially the issue of stratification.

Creating many small strata can potentially lead to problems, and it was suggested to use a top-down approach starting at a very coarse level with regards to NACE classification. Then division into subgroups can be used to create a suitable grouping. It was also pointed out that the same size classes are not necessarily relevant for all NACE groups. It was suggested to look further into the “cumulative square root f” method by Daelenius and Hodges in order to create optimum strata boundaries. This method is implemented in the R-package ‘stratification’ available on CRAN.

Level of detail is a trade-off between standard error on the total results and standard error on small strata. If a survey should report on a very detailed level, the sample must cover many small strata and hence the total will not be as well covered as in an optimally stratified sample.

Ad. 2. Due to the lack of data for a new SBS sample, a number of examples regarding estimation in the presence of nonresponse were prepared in continuation of the scripts for creating a RTS sample during mission 3.7.2. More specifically, three R scripts were prepared as described below:

- P07_RandomNR_20160212.R: The script applies a totally random nonresponse mechanism to the RTS sample, and it is shown how the right level of the estimates can be regained by inflating the original design weights by the inverse of the observed response propensities within each stratum. By means of simulation it is demonstrated that this simple correction does not introduce bias relative to the full sample provided that the nonresponse mechanism is truly random. However, this is rarely (if ever) the case in practical applications.
- P08_Calibration_20160212.R: An application of the calibration estimator implemented in the R-package 'survey' through the function `calibrate()` is demonstrated. The original sampling design did not include the variable describing region in any way. Hence, the number of units in each region estimated from the full sample using the HT-estimator deviates from the true and known number. Formulation of a calibration equation with estimated number of units within each region provides as a simple example of calibration. As was rightfully pointed out by the KAS staff, there is not much explanatory power in this variable, and hence no great gain in precision (reduced bias or reduced variance) can be expected from this.
- P09_CalibrationForNonresponse_20160215.R: The script applies a systematic nonresponse mechanism linked to the variable region. Thereafter calibration using region as auxiliary variable is applied. Through simulation it is shown, that some of the detrimental effects of systematic nonresponse are accounted for when using a proper auxiliary vector.

The R-files are made available for KAS, but please note that the scripts are prepared for demonstration purposes and should not be considered production grade code. Also, note that to run the code scripts and data from the previous mission 7.3.2 should reside in the same library as the provided script 7-9.

Ad. 3. An ODBC connection from SQL database to R was constructed. It was shown how functions can be defined in R, which can make it easy to extract data from SQL database to R and how to adjust these extractions to cover e.g. companies with given activity codes. It was also shown how graphs could be defined in R reading data directly from SQL database.

It was shown how data could be read directly into Excel pivot tables from SQL database, by ODBC connection and hence avoiding the step with saving data as .csv file or other alternative ways of exporting data from SQL database to Excel.

3. Conclusions and recommendations

It was demonstrated through examples how estimation in presence of nonresponse (either random or systematic) can be carried out in R. Also, a stratification strategy for the SBS survey was discussed.

A solution for communicating directly between a SQL database and the R software was established, and an effective way of communicating directly between SQL database and Excel was demonstrated.

The impression of the consultants is, that the staff at KAS are in general excited about the possibilities with R in statistical production. A more general course in R, as the one that will start ultimo February 2016 by provision of SIDA, will be a good start for use of R software in the statistical production. Another prerequisite for this is effective access to the data within the KAS data base structure. The connection from database to R established during this mission can also be very useful when doing reports and descriptive statistics directly on the data in the database.

Annex 1. Terms of Reference



EU Twinning Project
KS12 IB ST 01
Support to Statistics

Terms of Reference:**Component 3: Business Statistics****Activity 3.7.3: Methodology Training in Business Statistics****Scheduling:**

ToR –ready date: 25 January 2016
Start / end of activity: 9-12 February 2016
Reporting time: 19 February 2016

Mandatory result of the component:

Mandatory Result	Intervention logic	Benchmarks	Sources of information	Assumptions
Mandatory Result 3.7	Improved understanding of methodological issues	Staff from Business Statistics trained in various methodological issues relevant to their daily jobs	Quarterly reports from the Twinning Project Mission reports KAS website	<ul style="list-style-type: none"> • Senior management committed to training activities • Sufficient absorption capacity • Low turn-over of staff involved in implementation • Staff works on project related tasks in between missions • A detailed Terms of Reference is developed in a timely manner detailing tasks (input), expected output, participants of the activity and agenda

Subject / purpose of activity: 3.7.3 activity

This mission is a continuation of the first two missions on methodology 3.7.1 and 3.7.2.

The first missions were an assessment mission discussing the organization and state of the methodology work at KAS and it specified training needs within this field.

The second mission concentrated mainly on drawing a new sample for Retail Statistics and on introducing the software R.

The present mission will:

- Draw a new sample for SBS
- Discuss and train estimation in the presence of non-response. Using SBS as an example
- Discussing how to use “R” in connection with a database system

Expected output of activity 3.7.3:

Mission report describing the way ahead for methodology at KAS

Training on methodology issues conducted

New sample for SBS designed

KAS resources:

Mr. Ismajl Sahiti, Head of Division, ismail.sahiti@rks-gov.net

Mr. Bekim Canolli, Head of Methodology Division and Quality Manager; bekim.canolli@rks-gov.net

Ms. Servete Muriqi, Senior Officer in Methodology, servete.muriqi@rks-gov.net

Mr. Muhamet Kastrati, Manager of database, muhamet.kastrati@rks-gov.net

Mr. Hysni Elshani, Head of sector, hysni.elshani@rks-gov.net

Ms. Valdete Navakazi, high officer for STS, valdete.navakazi@rks-gov.net

Ms. Luljeta Krasniqi, high officer for SBS, luljeta.krasniqi@rks-gov.net

Mr. Bekim Bojku, high officer for energy statistics, bekim.bojku@rks-gov.net

Ms. Nazmije Belegu, high officer for transport statistics, naska1@hotmail.com

Ms. Sulltane Gashi, senior officer for tourism statistics, sulltanegashi@hotmail.com

Ms. Violeta Sylja, new staff, violeta.sylja@rks-gov.net

Ms. Saranda Berisha, new staff, saranda.h.berisha@rks-gov.net

KAS Twinning team:

Project Leader Mr. Ilir T. Berisha, Director of Economic Statistics and National Accounts,

Ilir.T.Berisha@rks-gov.net

RTA Counterpart Ms. Teuta Zyberi, International Relations Officer, teuta.zyberi@rks-gov.net

Member state resources:

Mr. Peter Tibert Stolze, Head of Division, Methodology and Analysis, Statistics Denmark, psl@dst.dk

Mr. Regin Reinert, Head of Section, Methodology and Analysis, Statistics Denmark, reg@dst.dk

Twining ressources:

Mr Per Knudsen, RTA, pkn@dst.dk

Ms Nora Zogaj, RTA Assistent, nzogaj@yahoo.com

Description of the background for the activity

In the first mission on overall assessment of Business Statistics in January 2014 (Ottosen and Netterstrøm) the consultants found that KAS has a relatively good foundation for business statistics.

Related to methodology work at KAS they found that it would be necessary to evaluate the current data collection methodology and questionnaire with a view to EU-regulations.

Furthermore, they say, it is important to develop sample updates, additional validation, data processing (indices), analysis and publication where Twinning experts should provide practical support to KAS. Especially methodology concerning non-response, the shift from NACE rev.1 to rev. 2 and the productions of indices is in demand.

It will be the recommendation that adjustments in the current methodology and collection will take into account the available resources and prioritize changes that compromise long term development and coherence.

Within the Twinning Project special focus is on development of SBS as these statistics are very important for the further development of National Accounts. The experts working with SBS at KAS found that administrative data should be used in the future to check the consistency of the data collected and to complement the surveyed enterprises data.

It is recommended that an IT tool for data editing should be developed and should include validation and checking rules.

Activities to be undertaken in preparation for the mission:

- All previous mission reports from component 3 of the Twinning Project can be found at www.dst.dk/kosovo

The expected activities are:

- Training on selected items of methodology
- Drawing a new sample for SBS
- Discussing how to use “R” in connection with a database system

Expected output:

- Mission report – according to template

Agenda, - February 2016

Day	Place	Time	Event
1	KAS	10:00-16:00	Introduction to this part of Component 3 Agreeing on the agenda “R” in connection with a database system
2	KAS	09:00-16:00	New sample for SBS. Using “R” Estimation in the presence of non-response “R” in connection with a database system
3	KAS	09:00-16:00	Estimation in the presence of non-response “R” in connection with a database system Conclusions and recommendations
4	KAS	09:00-11:00	Report writing. Debriefing: Experts, Component Leader and RTA

Annex 2. Persons met at KAS

Methodology Division:

Mr. Bekim Canolli, Head of Methodology Division and Quality Manager

Ms. Servete Muriqi, Senior Officer in Methodology

Business Statistics:

Mr. Ismajl Sahiti, Head of Division

Mr. Hysni Elshani, Head of sector

Ms. Valdete Navakazi, High officer for STS

Ms. Luljeta Krasniqi, High officer for SBS

Mr. Muhamet Kastrati, Database Manager of SBR

KAS Twinning team:

Project Leader Mr. Ilir T. Berisha, Director of Economic Statistics and National Accounts,

RTA Team:

Per Knudsen, RTA

Nora Zogaj, RTA Assistant Interpreter

DST Team:

Peter Tibert Stoltze

Regin Reinert