

**TWINNING CONTRACT**  
**Support to Statistics**

**Kosovo**



**MISSION REPORT**

**on**

**Methodology Training in Business Statistics**

**Component No. 3.7.2**

Mission carried out by  
Peter Tibert Stoltze and Regin Reinert,  
Statistics Denmark  
16-20 November 2015


*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

Executive Summary .....	<b>Fejl! Bogmærke er ikke defineret.</b>
1. General comments .....	4
2. Assessment and results .....	4
3. Conclusions and recommendations .....	6
Annex 1. Terms of Reference .....	6
Annex 2. Persons met .....	11

## 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 second mission to be devoted to Methodology Training in Business Statistics within Component 3 of the project. The mission was aimed at introducing the R software for statistical programming in general and for constructing survey samples in particular. A proposition of a sampling scheme for the RTS was presented.

The concrete objectives of the mission were:

- Introduce and discuss implementation of the statistical software R
- The MS experts should work with the BC professionals on drawing a new sample for Retail Statistics
- Use R for working with the new sample
- If there is time for it: Conduct training in Weighting for non-response and Calculation Sampling Errors

The consultants would like to express their 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 the work of the consultants.

**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 ToR stated 4 activities that should be

1. Introduce and discuss implementation of the statistical software R
2. Following a special specific request from KAS the MS experts will work with the BC professionals on drawing a new sample for Retail Statistics
3. If possible use R for working with the new sample
4. If there is time for it: Conduct training in Weighting for non-response and Calculation Sampling Errors

Ad. 1. R is an open source statistical software which can be very useful for KAS as a supplement to SPSS and Excel. On the first two days of the mission (November 16-17) the R software was introduced and installed on the laptops of delegates from KAS who joined the sessions. It was demonstrated how data can be read into R, by reading the population for the RTS into R. Finally some examples on how R can be used for examining data, cleaning data, enriching data, and transforming data were demonstrated.

Ad. 2. A table showing groups of retail businesses crossed with number of employees was constructed. The grouping of retail businesses was based on NACE codes. The 4 digits code was assessed to be too fine and that the 3 digits code too coarse. A suitable grouping of retail businesses would be somewhere between the 3-digits and the 4-digits NACE code. A proposal where some of the 3 digits groups were split in smaller groups was presented. The technique of constructing appropriate sampling groups was discussed.

Ad. 3. A sample was constructed based on the grouping suggested in point 2. It was shown how the package “sampling” in R can be used to perform simple random sampling (SRS) and stratified simple random sampling (STSI).

Ad. 4. The topic of weighting for non-response and calculating sampling errors is much wider than time on this mission allowed. However, an example of how this can be done was demonstrated again using the “sampling” package and a paper by Yves Tillé and Alina Matei<sup>1</sup>. If these topics should be covered at any depth it is recommended to schedule a mission devoted just for this purpose.

A number of R scripts were prepared for the RTS sample:

- P01\_ReadRetailData\_20151118.R: Read population data from .csv file. Examining and cleaning data, by removing possible duplicates and not used variables. Transformation of character variables to numeric variables and coding of missing values.
- P02\_InvestigatePopulation\_20151118.R: Investigation of suspicious values. Construction share distribution functions in order to gain insight into the population structure and possibly identify suspicious records.
- P03\_Stratification\_20151118.R: Enrichening population with regional codes and alternative activity codes. Construction of a strata table, based on activity and number of employees, that can be used for sampling.
- P04\_Allocation\_20151119.R: Creating a sample based on the table from program 3. Identifying vital parts of the population where a take-all sample should be applied. For the take-some strata an optimal allocation is calculated and certain manual corrections adapted.
- P05\_StratSampling\_20151119.R: Drawing a STSI sample from the population based on the allocation from program 5. The “sampling” package in R is used for drawing the sample.
- P06\_Estimation\_20151119.R: The “survey” package is used for calculating point estimates and standard errors of target variables. Calculation of domain estimates at the regional level is also done.

The R-files that are made available for KAS can be used as building bricks for construction other survey samples. However, further training in the use of R will probably be needed. A thorough course on estimation and calculation of sampling errors could broaden the use of R in KAS to other areas than just drawing the sampling and perform simple HT-estimation.

---

<sup>1</sup>Teaching Survey Sampling with the ‘sampling’ R Package  
[http://iase-web.org/documents/papers/icots8/ICOTS8\\_4J1\\_TILLE.pdf](http://iase-web.org/documents/papers/icots8/ICOTS8_4J1_TILLE.pdf)

### **3. Conclusions and recommendations**

It was concluded that R is very well suited for constructing samples in KAS. For smaller surveys it worked well with reading data from .csv files. For larger surveys a direct connection to the SQL database might be preferred.

This mission has demonstrated a few of the possibilities for handling data and performing statistical calculations with the R software. Generally, the software was very well received by the KAS staff, but obviously further training is needed in order to make R a viable tool in the daily production. This could be done either by formal training (a lecture format) or by allowing KAS staff to work with R in study groups and apply a learning-by-doing approach, possibly duplicating results from actual production.

In order to be able to apply more sophisticated estimation techniques utilizing the rich source of auxiliary data, another mission within this twinning project could be planned. Topics could include ratio estimation and regression estimation along with handling nonresponse.

## Annex 1. Terms of Reference

### Component 3: Business Statistics

#### Activity 3.7.2: Methodology Training in Business Statistics

**Scheduling:**

ToR –ready date: 1 November 2015  
 Start / end of activity: 16-20 November 2015  
 Reporting time: 27 November 2015

**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"> <li>Senior management committed to training activities</li> <li>Sufficient absorption capacity</li> <li>Low turn-over of staff involved in implementation</li> <li>Staff works on project related tasks in between missions</li> <li>A detailed Terms of Reference is developed in a timely manner detailing tasks (input), expected output, participants of the activity and agenda</li> </ul>

**Subject / purpose of activity: 3.7.2 activity**

This is a continuation of the first mission on methodology (3.7.1).

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

This mission will:

- Following a special specific request from KAS the MS experts will work with the BC professionals on drawing a new sample for Retail Statistics
- Introduce and discuss implementation of the statistical software R
- If possible use R for working with the new sample
- If there is time for it: Conduct training in Weighting for non-response and Calculation Sampling Errors

**Expected output of activity 3.7.2:**

Mission report describing the way ahead for methodology at KAS  
Training on methodology conducted  
New sample for Retail Statistics designed

**KAS resources:**

Mr. Ismajl Sahiti, Head of Division, [ismail.sahiti@rks-gov.net](mailto:ismail.sahiti@rks-gov.net)

Mr. Bekim Canolli, Head of Methodology Division and Quality Manager; [bekim.canolli@rks-gov.net](mailto:bekim.canolli@rks-gov.net)

Ms. Servete Muriqi, Senior Officer in Methodology, [servete.muriqi@rks-gov.net](mailto:servete.muriqi@rks-gov.net)

Mr. Muhamet Kastrati, Manager of database, [muhamet.kastrati@rks-gov.net](mailto:muhamet.kastrati@rks-gov.net)

Mr. Hysni Elshani, Head of sector, [hysni.elshani@rks-gov.net](mailto:hysni.elshani@rks-gov.net)

Ms. Valdete Navakazi, high officer for STS, [valdete.navakazi@rks-gov.net](mailto:valdete.navakazi@rks-gov.net)

Ms. Luljeta Krasniqi, high officer for SBS, [luljeta.krasniqi@rks-gov.net](mailto:luljeta.krasniqi@rks-gov.net)

Mr. Bekim Bojku, high officer for energy statistics, [bekim.bojku@rks-gov.net](mailto:bekim.bojku@rks-gov.net)

Ms. Nazmije Belegu, high officer for transport statistics, [naskal@hotmail.com](mailto:naskal@hotmail.com)

Ms. Sulltane Gashi, senior officer for tourism statistics, [sulltanegashi@hotmail.com](mailto:sulltanegashi@hotmail.com)

Ms. Violeta Sylla, new staff, [violeta.sylla@rks-gov.net](mailto:violeta.sylla@rks-gov.net)

Ms. Saranda Berisha, new staff, [saranda.h.berisha@rks-gov.net](mailto: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](mailto:Ilir.T.Berisha@rks-gov.net)

RTA Counterpart Ms. Teuta Zyberi, International Relations Officer, [teuta.zyberi@rks-gov.net](mailto:teuta.zyberi@rks-gov.net)

**Member state resources:**

Mr. Peter Tibert Stolze, Head of Division, Methodology and Analysis, Statistics Denmark, [psl@dst.dk](mailto:psl@dst.dk)

Mr. Regin Reinert, Head of Section, Methodology and Analysis, Statistics Denmark, [reg@dst.dk](mailto:reg@dst.dk)

**Twinning resources:**

Mr Per Knudsen, RTA, [pkn@dst.dk](mailto:pkn@dst.dk)

Ms Nora Zogaj, RTA Assistent, [nzogaj@yahoo.com](mailto: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](http://www.dst.dk/kosovo)

**The expected activities are:**

- Training on selected items of methodology
- Introduce the statistical software R

**Expected output:**

- Mission report – according to template

**Program, - November 2015**

<b>Day</b>	<b>Place</b>	<b>Time</b>	<b>Event</b>
<b>1</b>	KAS	10:00-16:00	Introduction to this part of Component 3 KAS: Actual state of methodology at KAS Discussion on the agenda New sample for Retail. Who will do what? Introduction and discussion of R (SIDA will be invited also)
<b>2</b>	KAS	09:00-16:00	New sample for retail
<b>3</b>	KAS	09:00-16:00	New sample for retail continued. Training (if time permits): Calculation of sampling errors and Weighting for non-response
<b>4</b>	KAS	09:00-16:00	New sample for retail continued Any other business Report writing
<b>5</b>	KAS	09:00-12:00	Conclusions, 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. 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