TWINNING CONTRACT

JO/13/ENP/ST/23

Strengthening the capabilities of the Department of Statistics in Jordan

MISSION REPORT

on

Activity 2.1: Assessment of current use of sampling methodologies

Mission carried out by

Mr Peter Stoltze, Statistics Denmark
Mr Mārtiņš Liberts, Central Statistical Bureau of Latvia

8th to 12th December 2013

Version: Final

STATISTICS DENMARK

Istat
Istituto Nazionale di Statistica

DOS 1949

LATVIJAS STATISTIKA

Expert contact information

Peter Stoltze
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 39173827
Email: psl@dst.dk

Mārtiņš Liberts
Central Statistical Bureau of Latvia
Lāčplēša Street 1,
Riga, LV-1301,
Latvia
Tel: +371 67 366 833
Email: martins.liberts@csb.gov.lv

Table of Contents

List of Abbreviations

DoS	Department of Statistics of Jordan
EU	European Union
EUS	Employment and Unemployment Survey
FD	Frames Division
ICT	Internet and Communication Survey
IES	Income and Expenditure Survey
JCS	Job Creation Survey
MSTD	Methodologies and Statistical Techniques Directorate
PFHS	Population and Family Health Survey
SD	Sampling Division
SPSS	IBM SPSS Statistics
RTA	Resident twinning adviser
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 first mission to be devoted to Component 2 of the project.

The purposes of the mission were to assess the current situation at DoS regarding survey sampling (including sample design, sample allocation, actual sampling, imputation, weights, and sample error calculation) and seasonal adjustment.

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 consultant and do not necessarily correspond to the views of EU, DoS or Statistics Denmark or Central Statistical Bureau of Latvia.

2. Assessment and Results

During the stay we have met with three directorates:

- Methodologies and Statistical Techniques Directorate (household surveys);
- Directorate of Economic Surveys;
- Directorate of Agricultural Statistics.

All employees have been very kind and provided us with the requested information. The following is a brief overview of information gained during the mission.

Overview of current use of sampling methodologies and data collection sources
All applied sampling designs are sound and adapted to the actual conditions under which the surveys are carried out. Most of the frames are based on a census (which may be several years old), that is being partially updated during the field work. Also, the frames have very little auxiliary information due to the limited cooperation with other governmental bodies. Rich sampling frames are a prerequisite for vastly improved sampling schemes.

Data collection is carried out by personal interview, either paper assisted (PAPI) or computer assisted (CAPI). The use of web technology or even phone for data collection seems far away due to lack of suitable contact information. The interviewing using phone is not feasible also because respondents in general are reluctant to disclose personal information over phone. Personal interviewing is much more preferred regarding this aspect.

Overview of needs for various sampling techniques
Especially for the economic surveys, the concept of *domain estimates* should be better utilized. Domain estimation allows design based (or model assisted) estimation in arbitrarily defined domains as long the domain contains at least one observation (however precision has to be monitored carefully and disclosure control has to be insured in case of domains with small number of respondents). Planned tabulation could go into the definition of strata, but strictly speaking this is not necessary.

Overview of calculation of sampling error
For the household surveys, there is some effort to calculate sampling error (with the SPSS software and IMPS (Integrated Microcomputer Processing System)). These calculations were not discussed in any details. Other surveys did not calculate sampling errors.

The calculation of sampling errors for composite designs often involves complicated computations making the use of suitable software important. There was an expressed wish to continue using SPSS, which only to some degree can facilitate this. We recommended to look further into the use of R (free software with many useful functions for survey sampling), but the DoS staff were (rightfully) hesitant to make such a commitment. Change of core software is a high-level strategic decision, which involves a significant investment even if the software is free. However, R can be considered as an alternative tool in some cases to increase the flexibility of DoS to adapt to different environments.

Overview of the calculation of sampling weights in different surveys

Most of the surveys presented are based on some kind of single-stage or multi-stage sampling, and the total sampling weight is calculated as the product of the inverse inclusion probabilities at the different stages. This is the proper way to calculate weight when there is no non-response. However, non-response does occur, and it is handled by relatively simple weight adjustment. This practice has severe bias potential since the systematic differences that might characterize the non-respondents from the respondents are ignored. A better way to deal with non-response is through weight calibration. The principles were outlined, and a document containing theory on this topic ("Estimation in the presence of Nonresponse and Frame Imperfections" from Statistics Sweden) was handed over in electronic form. The theory in this document could very well be the core part of one of the training activities during this component.

Draft training plan for the needed competence development on statistical methods

The training plan was drafted based on the information gained during the meetings with DoS staff. The training plan was discussed and adapted in meetings with Mrs Batoul Obaid, Mr Ahmad Mowafi and Mr AbedWadood Matouk. The proposed training plan consists of the following points:

1. Training on statistical software R. *Microsoft Excel* is used quite extensively in different stages of survey sampling currently. This could be a weak point in the statistical production chain.
2. Training on imputation (for household surveys). It is used only in Income and Expenditure Survey (IES) currently. There is no need for more imputation now as item non-response is low. But there is a willingness to have more knowledge about it and willingness of being ready when response rates will decline in the future.
3. General training on compiling economic statistics (sampling, sample size allocation, estimation, weight calibration, domain estimation). Needed because Sampling Division (SD) does not have enough experience with economic surveys. The training would be useful also for Economic Directorate (there was an observation of lacking knowledge about domain estimation).
4. Theoretical and practical training on weight calibration. Calibration currently is done with Microsoft Excel. This is an obstacle to include more auxiliary variables in a calibration model. It is highly recommended to reconsider the use of R for this task.
5. Theoretical and practical training on sampling error estimation. Again it is highly recommended to reconsider the use of R for this task.
6. General training on register assisted statistics, the concepts about statistical registers, building and maintaining statistical registers.
7. There were not any observations made about seasonal adjustment of time series in household, economic or agriculture surveys. The course on seasonal adjustment has to be tailored for audience without practical experience in the topic.
8. Agriculture Directorate is very keen on implementing new surveys, methods and techniques (according to the National Agricultural Statistics Development Strategy in Jordan, January 2013). There is a specific need for assistance to arrange transfer from census to sample survey for the Cultivated Area Survey in Jordan Valley (*Aghwar*).

Transfer of the Danish, Latvian and the European Union experience in sampling methodologies

On Monday afternoon and Wednesday morning, we gave presentations (slides in English – available upon request) in the conference room. Wednesday morning included a demonstration of calibration and standard error calculation in R. (Memory stick with programs and relevant tutorials were handed over, but inter-package dependencies means that there will still be need of internet access.) There was

limited participation (3-4 employees from DoS), and the time allotted was shortened due to other meetings. In total, 2+2 hours were spent on presentations.

Seasonal adjustment
To the best of our knowledge no seasonal adjustment is being conducted at DoS. Thus, the training activity concerning seasonal adjustment must take this into account and be introductory. It was suggested, that the training course introduces the Demetra+ (or perhaps JDemetra+) software provided by Eurostat. This software facilitates both of the currently recommended methods for seasonal adjustment, namely X-12-ARIMA (US Census Bureau) and TRAMO/SEATS (Bank of Spain).

Description and assessment of the situation
Basis for conclusions and recommendations in section 3 below

The sampling techniques currently being applied for sample surveys are reasonable when taking into account the available sampling frames and survey modes used.

The weighting and estimation can be improved by introducing weight calibration as a common weighting procedure. The standard software for weight calibration has to be considered and introduced in the statistical production process. The modernisation of weighting procedures will allow usage of more auxiliary information for weighting. However, if no efforts are made towards more advanced sampling frames (i.e. with more auxiliary information) this might be of little value.

The imputation techniques are not widely used as response rates are quite high for most of the surveys. The knowledge about imputation can be improved to increase the capability of DoS to deal with future situations when response rates might decrease.

The calculation of sampling errors has to be introduced as a standard procedure for most of the sample surveys. The standard software for sample error calculation has to be considered.

The application of seasonal adjustment was not detected during the mission. The knowledge about seasonal adjustment has to be improved.

Achievements of the mission
Assessment and review of sampling methodologies, calculation of sampling errors and sampling weights in different surveys has been done during the mission. The results of the assessment are available in the mission report.

Areas of improvement have been detected during the mission. The training plan has been drafted based on the observed areas of improvement. The draft training plan has been discussed with the project leader during the mission. The draft training plan is included in the mission report.

Note: The programme on Thursday was cancelled due to the weather situation[1].

3. Conclusions and Recommendations

The visit has provided us with information about the current state of affairs, which is a good basis for drafting a training plan for future activities.

Action	Deadline	Responsible person
Mission Report Prepared	January 2014	Peter Stoltze, Mārtiņš Liberts
Adapt the next missions of the component according to the draft training plan (if necessary and possible)	January 2014	Thomas Olsen

[1] http://jordantimes.com/govt-announces-thursday-public-holiday-due-to-weather-conditions

Annex 1. Terms of Reference

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 2014)
- 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

- Assessment and review on the current use of sampling methodologies in different surveys
- Priority to be given to the topics in the component
  - Design of surveys
  - Seasonal adjustment
  - Weighting, imputation, non-response and sampling errors
- Identification of areas to improve
- Assessment and review on the calculation of sampling error
- Assessment and review on the calculation of sampling weights in different surveys
- Identify training needs and draft a training plan on how to ensure the competence development of staff

2. Expected output of the activity

- Overview of current use of sampling methodologies and data collection sources
- Overview of needs for various sampling techniques.
- Overview of calculation of sampling error.
- Overview of the calculation of sampling weights in different surveys
- Draft training plan for the needed competence development on statistical methods
- Transfer of the Danish and Latvian, and in general the European Union, experience in sampling methodologies

3. Participants

DoS

Mrs Batoul Obaid, Director of Methodologies and Statistical Techniques Directorate *(Component Leader*)
Mrs Ghaida Khasawneh, Head of Sampling Division
Mr Mohammad Thnyyan, Head of Frames Division
Mr Abd Alnaser Aljarere, Statistician in Sampling Division
Ms Wafaa Amer, Statistician Assistant in sampling Division
Mr Ayman AlQasem, Statistician in Frames Division

MS experts

Mr Peter Stoltze, Deputy Head of Research and Methods Division, Statistics Denmark
Mr Mārtiņš Liberts, Deputy Head of Mathematical Support Division, Central Statistical Bureau of Latvia

Other stakeholders taking part in the activity

Staff from other relevant DoS Directorates will be invited.

Programme for the mission

Time	Place	Event	Purpose / detail
Sunday, morning	Hotel / DoS	Meeting with RTA	To discuss the programme of the week
Sunday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussion on the current use of sampling methodologies in household surveys
Sunday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Discussion on the current use of sampling methodologies in economic surveys
Monday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussion on the current use of sampling methodologies in agriculture surveys
Monday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Presentations about the Danish and Latvian (or other country) experiences on the sampling methodologies and organization of work.
Tuesday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussion on the current use of sampling errors for the households surveys
Tuesday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Discussion on the calculation of sampling errors for the economic and agriculture surveys
Wednesday, morning	DoS	Meeting with BC Component Leader and BC Experts	Presentations about the Danish and Latvia (or other country)experiences on the sampling errors
Wednesday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Discussion and evaluation of current use of the sampling weight
Thursday, morning	DoS	Meeting with BC Component Leader	Presentation of MS Experts' findings and agreement on the reached conclusions
Thursday, morning	DoS	Ad-hoc meetings	Final clarifications with BC Experts, preparation of report and presentation for BC Project Leader
Thursday, noon	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:

Mr AbedWadood Matouk, project leader
Mr Ahmad Mowafi, RTA counterpart

Methodologies and Statistical Techniques Directorate:
- Mrs Batoul Obaid, Director (component leader)
- Mrs Ghaida Khasawneh, Head of Sampling Division
- Mr Mohammad Thnyyan, Head of Frames Division
- Mr Abd Alnaser Aljarere, Statistician in Sampling Division
- Ms Wafaa Amer, Statistician Assistant in sampling Division
- Mr Ayman AlQasem, Statistician in Frames Division

Economic Directorate:
- Mr Nimer Hashem Gharbia, Head of Industry and Energy Division

Agricultural Directorate:
- Mr Bassam Zain, Head of Crops Production Division
- Mr Sae'd Shawawreh
- Mr Haitham Ammari
- Mr Abdelnaser Obeidat
- Mr Mohammad Al-Jawarneh

RTA Team:

Thomas Olsen, resident twinning adviser (RTA)
Amal Aliah, RTA assistant
Deena Moghrabi, interpreter

Annex 3. Notes on the Current Use of Sampling Techniques

Overview

Methodologies and Statistical Techniques Directorate (MSTD) was developed at 2009. There are two divisions:

- Sampling Division (SD);
- Frames Division (FD).

Sample surveys are split in three groups:

- Household surveys;
- Economic (business) surveys;
- Agriculture surveys.

The main task for MSTD is to support the production of household statistics (sampling, imputation, weighting, and sampling error estimation). It is planned to support economic and agriculture surveys in the future (currently this is done within the Economic Directorate and the Agricultural Directorate, respectively).

Household Surveys

Work Organisation

The application of survey sampling methodologies is organized by two divisions – SD and FD. The first stage sample (area sample) is drawn by SD. The area sample is delivered to FD. FD does the household frame update in each sampled area (quarterly). The second stage sample (household sample) is drawn by SD from the updated household frame in the sampled areas.

SPSS and Microsoft Excel are used for sampling. The choice of the software depends on an employee (on his or her skills).

Surveys

There are four regular surveys:

- Employment and Unemployment Survey (EUS) – quarterly (rotating quarterly panel is used);
- Job Creation Survey (JCS) – biannual;
- Internet and Communication Survey (ICT) – annual;
- Income and Expenditure Survey (IES) – every two years (there is a quarterly panel used).

There are other surveys, for example Population and Family Health Survey (PFHS). The master sample is used for EUS, JCS and ICT (the term “master sample” is described further below). EUS is a sub-sample of JCS and ICT is a sub-sample of EUS. Samples for EIS and PFHS are independent.

Sampling Frames

The information from the last population census (in 2004) is used to build the area frame. The population census is planned to be done once every 10 years. The next population census was planned for 2014, but it is postponed till 2015 because of the Syrian crises. It could be postponed even longer if the numbers of refugees will not decrease noticeably.

The frame of households is partly updated every quarter. The update of household frame is done in the sampled areas selected for the quarterly sample. The update of the household frame is done in the field by interviewers. Interviewer draws a map of an area, lists all households in that area, and then records the name of the head of each household. There are instructions (in Arabic) available for interviewers about how to do the frame update. Data collected during the frame update are recorded in a database. The full update of the household frame is done only during the population census. More details about

the sampling frame update are available from the attachment "DoS Procedure for Frame update.pdf" provided by DoS.

The average size of a cluster (area) is 72 households (according to the data of the last census). The cluster size is ranging from 30 to 192 households. The question was raised how to deal with situation when the size of clusters have changed dramatically.

Sampling Design

Two-stage sampling is used for household surveys. The first stage sampling units are census counting areas from the last population census (the last population census was in 2004). There are around 14,000 areas in the area frame. The areas are stratified by geographical, social and economic variables (for example – owning of durable goods, owning of property, there are 17 variables used for stratification). There are 30 strata.

The areas are sampled by systematic πps sampling (sampling with probabilities proportional to area size). The number of households from the previous population census is used as a size measure for areas.

Households are second stage sampling units. Ten households are sampled in each cluster for EUS, ITC and EIS, 20 households are sampled in each cluster for PFHS and 30 households are sampled in each cluster for JCS.

Survey Mode and Response

Usually all individuals from a sampled household are sampled (in EUS, JCS, and ICT). Kish tables are used to sample individuals from a sampled household if only part of individuals have to be surveyed.

The only mode used for interviewing is face to face (CAPI or PAPI). Telephone surveys are not used because people hesitate to provide personal information over phone to unknown individuals. There are instructions for interviewers how to do the surveys. The average response rate is 92%.

Part of the respondents is re-contacted by the DoS and small questionnaire is asked. This is done in order to check the interviewers work.

Master Sample

A master sample is used for three surveys – EUS, JCS, and ICT. It is an area sample drawn for five years. The master sample is divided into 40 replicas (two new replicas are drawn every quarter for five years, $2 \bullet 4 \bullet 5 = 40$). Eight replicas are used every quarter, where four replicas are repeated from the previous quarter, two replicas are new, and the remaining two replicas are repeated from older quarter. The rotation of areas is introduced for EUS which is a panel survey with four waves (one household is surveyed for four quarters in a row). There were no problems or challenges mentioned about the master sample and three surveys using the master sample.

Employment and Unemployment Survey (EUS)

Description about EUS is available from the attached file "DoS_EUS_Background_2013.pdf" provided by DoS.

Income and Expenditure Survey (IES)

This is a panel survey where sampled households are monitored quarterly during an entire year. Expenditure data are collected quarterly but income data are collected twice a year. Different weeks are observed in each quarter.

Poverty statistics are produced from the IES. Quarterly and annual statistics are produced, but all statistics are published only annually.

The problem of IES is panel attrition. Panel attrition is compensated by unit or item imputation. A pilot study was done by surveying the respondent only once a year. The aim was to remove the attrition effect by dropping the panel component. This is possible because the panel component is not mandatory to derive quarterly and annual estimates. However, the results of the pilot study were not accepted by the Household Expenditure & Income Division who would like to stick to the current methodology with panel component (where respondent is surveyed four times during a year).

Population and Family Health Survey (PFHS)

PFHS is described in details by the publication of DoS and ICF International (2013, *Jordan Population and Family Health Survey 2012*, available electronically in Arabic and English). The results of the PFHS 2012 were presented in a seminar on December 10, 2013, where the consultants were invited to participate during the last sessions. There are no problems or challenges mentioned about PFHS.

Weighting

A weighting is done for all surveys at the household or individual level. The choice of weighting technique depends on the survey aims. Demographic information is used for weight adjustment.

Usual scheme of weighting is to compute design weights according to the sampling design and apply non-response adjustment in each cluster. For some surveys additional weight adjustment (calibration) using external demographic information is done.

Some examples of weighting observed during the meetings are given:

- Weight calibration is not applied for the EUS. It is because DoS is publishing only rates (totals are not published, totals are published by the Centre of Human Resources calculated using the EUS data collected by DoS);
- Combined – household and individual weighting is done for the ITC survey. However household and individual weights are computed independently. The coherence between household and individual estimates is not set as a requirement for the survey;
- The number of Jordanians and non-Jordanians is used as auxiliary information for weight calibration usually. The usage of more auxiliary variables in the calibration is prohibited usually because *Microsoft Excel* is used for weight calibration;
- Several auxiliary variables were used for the Child Labour Survey, for example – age group, gender, governorate (there are 12 governorates in Jordan). The totals of auxiliary variables were provided by the Directorate of Population & Social Statistics.

Imputation

Item non-response is low in household surveys in general. The missing answers are shown in tables as "Not known". Imputation is used for IES to reduce the attrition effect. Imputation methods used are: mean imputation, hot-deck and cold-deck (historical information with trend adjustment).

Variance Estimation

Variance estimation usually is done with *IBM SPSS Statistics* using the *IBM SPSS Complex Samples* module. The module covers the sampling designs used for the household surveys. However, the module does not cover weight calibration, so the effect of weight calibration is not covered by the variance estimates.

Economic Surveys

Work Organisation

Sampling for economic surveys is done by the Directorate of Economic Surveys. There are six divisions in the Directorate:

- Industry & Energy Division;
- Internal Trade Division;
- Transport, telecom & Tourism Division;
- Construction Division;
- Service & Finance Division;
- Employment Division.

Sampling Frames

Sampling frames for surveys are built from the data of the last economic census. The economic census for manufacture enterprises was done from 1959 till 1994. The economic census of all enterprises in all activities was done in 1999, 2006 and 2010. The economic census is done on field by interviewers with area maps. Buildings are visited and two questionnaires are used to detect economic units – short questionnaire to detect activity and long questionnaire for active economic units.

The sampling frame is updated by DoS. Large economic units are added to the sampling frame between censuses. Small economic units are added only from the census data.

There has been an attempt to develop a nation-wide enterprise register. Technical committee was set up for it. The committee consisted of the members from the DoS, Ministry of Industry and Trade, Municipalities and other governmental bodies. The first task achieved was the implementation of unique ID number for each unit. The work is on-going.

Sampling Design

There are two types of sampling designs used for the economic surveys. The most common sampling design is stratified simple random sampling.

Stratified systematic πps sampling is used for the manufacture surveys. The number of employees is used as a size measure for the systematic πps sampling design. The choice of the design is justified by the high correlation between the size variable (number of employees) and the study variable (the manufacture output). The size groups of units are defined by the number of employees – 1-4, 5-19, 20 and more. The census is used for the last size group. Sampling is used for the first two size groups. Different questionnaires are used for enterprises with 1-4 and 5-19 employees.

Microsoft Excel is used for sample selection.

More details about the sampling designs used for economic surveys are available from the attachment "DoS Economic surveys sample withdrawal.pdf" provided by DoS.

Other Notes

The overall non-response rate for economic surveys is from 10%. No variance estimation is done for any of the economic surveys.

Agriculture Surveys

Work Organisation

The sampling for agriculture surveys is done by the Directorate of Agricultural Statistics. There are three divisions at the directorate:

- Crops Production Division;
- LiveStock Production Division;
- Agricultural Economy Division.

National Agricultural Statistics Development Strategy in Jordan (January 2013) is used to plan the development of the agriculture statistics.

Sampling

The last farm census was done in 2007. The data of the last census are used for building sampling frames for most of the surveys. The description about sampling of main agricultural statistics surveys is available from the attachment "DoS_Sampling_of_Main_Agricult_Surv_2013.pdf" provided by DoS.

Other Notes

The response rate is very high for agriculture surveys, in most cases 100% or at least very close. The agriculture production is not taxable in Jordan and farm owners are very cooperative – these are the main reasons for such a high response rate.

The weighting for sample surveys is done according to the sampling design. There is not any need for non-response adjustment as full response is achieved in most cases. Imputation is not applied as the item-response rate is very high as well.

The data of agriculture surveys is stored in *Oracle* database and *Microsoft Excel* is used as tool for statistical work.