TWINNING CONTRACT

JO 21 ENI ST 01 22

Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices

MISSION REPORT

on

Component 2
Methodology for producing Small Area Statistics

Activity: 2.1.2:
Theory and best practice of Small area estimations

Mission carried out by
Dr. Danila Filipponi
Dr. Andreas Berg

Amman, Jordan
26-29 February 2024

Version: Template

Strengthening the capacity of Jordan's Department of Statistics

Authors' names, addresses, e-mails

Ms. Danila Filipponi
Senior Researcher, Directorate for Methodology and Statistical Process Design.
Head of the unit: Labour Register,
The Italian National Institute of Statistics (ISTAT)
Via C. Balbo, 15
00184 Roma
Italy
Email: dafilipp@istat.it

Dr. Andreas Berg
Statistician
The Federal Statistical Office in Germany (Destatis)
Gustav-Stresemann-Ring 11
65819 Wiesbaden
Germany
Email: Andreas.Berg@destatis.de

Strengthening the capacity of Jordan's Department of Statistics

Table of contents

Strengthening the capacity of Jordan's Department of Statistics

List of Abbreviations

- BC – Beneficiary Country
- DoS – Department of Statistics
- MS – Member State
- PL – Project Leader
- RTA – Resident Twinning Advisor
- SAE – Small Area Estimation
- STE – Short-term Expert

Strengthening the capacity of Jordan's Department of Statistics

Executive Summary

The mission was designed to equip DOS participants with a comprehensive overview of small area estimation (SAE) methodologies. Encompassing various SAE methodologies, the program covered requirements, estimation methods—both design-based and model-based—and techniques for assessing accuracy and model validity. Furthermore, the presentation facilitated a comparative analysis of different estimation approaches, emphasizing the importance of accuracy measures and adherence to international quality standards.

Each concept was elucidated through practical examples, demonstrating their real-world application and the evaluation of accuracy measures. Hands-on demonstrations utilizing R code applications, carried out with synthetic data, provided participants with invaluable insights into implementation strategies.

As a result, DOS participants emerged from the mission equipped with the basic knowledge and tools to effectively initiate the integration of SAE methodologies. Additionally, the mission aimed to pinpoint practical applications within DOS. To this end, a thorough review of the data landscape was conducted, identifying target indicators such as average household income, domains of interest like sub-districts, and pertinent auxiliary information crucial for SAE production within DOS, such as the 2015 census and civil register.

Strengthening the capacity of Jordan's Department of Statistics

1. General comments

This mission report was prepared within the Twinning Project "Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices". The current Mission is carried out as part of component 2 Methodology for producing Small Area Statistics. The actions planned for this activity were carried out as scheduled.

The purpose of this activity is to give a theoretical introduction small area estimation with focus on basic smoothing and data manipulation. The theory will be exemplified by practical exercises using artificial data provided by MS. The following subject were covered:

- Why is small area statistics needed
- Definition and theory behind SAE methodology
- Identification of the informative gaps with respect to the use small area methodology – both in respect to practical and technical side of small area statistics.
- Clarification of the needs
- Production of direct estimate and their variances, Production of other indirect estimators (synthetic and composite) from designs-based perspective
- Introduction to area and unit level model based SAE
- Discusion of indicators, data availability, and data preparation
- Outline a custom-fit design/model to be used in the next mission (With DoS data)

The consultants would like to express their sincere 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 their work. The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of the EU, ISTAT or Destatis.

2. Assessment and results

The main goal of this mission was to provide a theoretical introduction to the tools and components essential for small area estimation. The potential need, utility and requirements of small area estimation were thoroughly discussed. Given the wide range of available estimation methods encompassed by small area estimation, several classes of estimators were presented. Participants from DOS were introduced to classical design-based estimators, including Horvitz-Thompson, difference, ratio, post-stratified and regression estimators. The mission also addressed the prerequisites, advantages, and drawbacks associated with implementing these estimators, as well as appropriate accuracy measures. Moreover, international standards for establishing accuracy benchmarks were provided to serve as a reference point for evaluation.

Moreover, during the mission were introduced concepts and provided several examples of indirect estimators. It was covered the topic of synthetic estimators and their application examples at aggregated levels, through spatial combinations, and post-stratification, exemplifying the impact of bias and instability on accuracy measures and providing insightful illustrations for better comprehension. Additionally it was presented the topic of composite estimators, offering various examples and elucidating weighting concepts tailored to different components of this estimator.

Strengthening the capacity of Jordan's Department of Statistics

The participants from the DOS have been thoroughly prepared with an understanding of model-based estimation. Beginning with fundamental regression analysis techniques, were introduced different estimation levels depending upon the availability of auxiliary data. Specifically, it introduced concepts of unit-level modelling and area-level modelling. Additionally, the mission presented the rationale behind incorporating random factors at the domain level. Basic estimators for parameters within resulting unit- and area-level models were introduced, according to methodologies such as Battese-Harter-Fuller and Fay-Harriot

The discussions highlighted the importance of model checking, emphasizing the need to continually assess model assumptions, ascertain the relevance of auxiliary variables, and ensure the stability and reliability of final estimates. A particular emphasis was placed on the quality requirements for auxiliary information, which were explored in detail.

An extensive deliberations took place regarding the comparison of results and the selection of appropriate estimators from a diverse sets of options. Each approach was thoroughly examined, highlighting key distinctions, especially in terms of accuracy measure between design-based and model-based methodologies. Additionally, the pro and limitations of each approach were carefully presented, providing valuable insights for decision.

The statistical methodologies introduced were demonstrated through practical R code applications, offering valuable insights into their implementation. These applications were conducted using synthetic data. Synthetic data are artificially generated datasets designed to replicate the patterns and characteristics of real data. This process involves employing statistical models and algorithms while preserving the fundamental features of the original dataset. In the context of this course, synthetic data refer to datasets generated to mirror the Austrian EUSILC (European Union Statistics on Income and Living Conditions) survey data. These datasets serve as invaluable tools for research and analysis within the R environment, enabling users to explore and understand various methodologies with confidence. A detailed overview of the structure and contents of the dataset has been provided, outlining the composition of the data files, encompassing both unit-level data and aggregated data at the domain level. Then, throughout the course, the dataset has been extensively utilized in all applications, providing learners with a comprehensive understanding of the methodologies discussed. The R code, complete with detailed comments, has been shared with the DOS.

The theoretical and practical explanation of small area models, coupled with experience working with synthetic data developed during the course, generated a beneficial discussions among the participants. These discussions facilitated a deeper understanding of the key concepts, steps, and requirements involved in applying small area models.

On the final day of the mission, the focus shifted from theoretical discussions on small area estimation to practical applications. DOS experts led the session, presenting Stata programs used in the former World Bank Project on SAE models. These programs, employing the ELL methodology, aimed to generate small area estimations of poverty rates at district and sub-district levels, drawing from HEIS data. However, DOS experts highlighted difficulties encountered when attempting to run the Stata programs to produce updated SAEs.

Subsequently, in collaboration with the DOS experts, a comprehensive review of the data landscape was conducted to pinpoint the target indicator, domains of interest, and potential auxiliary information essential for SAE production in DOS. The target indicator identified was average Household Expenditure, while the geographical domains focused on sub-districts. Furthermore, as auxiliary information, Census data and Civil register data were identified.

Strengthening the capacity of Jordan's Department of Statistics

3. Conclusions and recommendations

The mission has provided a foundation for DOS experts to move into the practical application of SAE methodologies. Through a combination of theoretical discussions, hands-on training with R code applications, and collaboration with experts, participants have gained valuable insights into the complexities of SAE and its potential applications.

Moving forward, it is important to capitalize on this knowledge and continue building upon the skills acquired during the mission. This can be achieved by enhancing expertise in utilizing R software, drawing from the R course attended by DOS experts, and further familiarizing themselves with the small area estimation programs shared during the mission. Practical application of SAE models to DOS indicators will be crucial in solidifying the understanding and capabilities in producing reliable small area estimations.

The next mission will focus on the production of SAE on the identified target indicators, domains of interest, and potential auxiliary information. To achieve this goal, the following actions need to be taken:

Action	Deadline	Responsible person
Ensure Availability of HEIS data with sampling weights	21-25 April 2024	DOS experts
Perform Direct Estimation of Average Expenditure of Households and Standard Errors by Governorate, Districts, and Sub-districts	21-25 April 2024	DOS experts
Identify possible covariate in 2015 census data	10 April 2024	Danila Filipponi and Andreas Berg
Validate Identified Covariates and Generate Aggregated Means by Sub-districts	21-25 April 2024	DOS experts
Ensure Availability of Civil Register Data	21-25 April 2024	DOS experts

Strengthening the capacity of Jordan's Department of Statistics

Annex 1. Terms of Reference

Terms of Reference

EU Twinning Project JO 21 ENI ST 01 22

Component 2:
Methodology for producing Small Area Statistics

Activity 2.1.2:
Theory and best practice of Small area estimations – Part 1 Basic smoothing and data manipulation

Dates: 26-29 February 2024

Strengthening the capacity of Jordan's Department of Statistics

List of abbreviations

BC   Beneficiary Country

DoS   Department of Statistics

ESS   European Statistical System

MS   Member State

RTA   Resident Twinning Advisor

SAE   Small Area Estimation

SAS   Small Area Statistics

STE   Short Term Expert

ToR   Term of References

Strengthening the capacity of Jordan's Department of Statistics

0. Objective and Mandatory Results for the component

Objective
To review and develop the source data and methodology for producing Small Area Statistics.

Mandatory results and indicators for achievement for each sub-component

Table 1: Mandatory results and indicators for achievement for each sub-components within Component 2: Methodology for producing Small Area Statistics (SAS). Please be aware that despite not being explicit defined in the current mandatory results this current Missions will focus focusing on producing SAS on expenditure.

MR from the Twinning Fiche	Indicators
MR 2.1: Pilot project to assess inclusion of administrative and other external data sources in the development of SAS and action plan developed.	**Indicators 2.1.A:** Administrative and other data sources investigated and their potential assessed **Indicators 2.1.B:** Action Plan for inclusion of administrative data prepared **Indicators 2.1.C:** Technical infrastructure for transfer of administrative data developed
MR 2.2: Develop methodology for producing SAS on poverty and provide recommendations on how this methodology can be applied to other areas within the DoS.	**Indicators 2.2.A:** Methodology proposed including the potential use of modelling techniques, building on work in 2.1 above **Indicators 2.2.B:** Analysis completed on how new methodology can be expanded to other statistical areas
MR 2.3: Develop training programs and manuals for use in the DoS and partner institutions based on pilot project outcomes.	**Indicators 2.3.A:** Detailed documentation on statistical standards, classifications, identifiers, etc. developed **Indicators 2.3.B:** Comprehensive training programme and workshops provided for DoS staff and partner institutions **Indicators 2.3.C:** DoS leadership role in ensuring proper statistical standards applied across the Jordanian statistical system reinforced
MR 2.4: Implement communication strategy with stakeholders on strengthening small area statistics.	**Indicators 2.4.A:** Communications strategy in place for users and data providers on importance of SAS **Indicators 2.4.B:** New statistical outputs and greater media exposure, including via social media

2. Purpose of the activity
The purpose of this activity is to give a theoretical introduction small area estimation with focus on basic smoothing and data manipulation. The theory will be exemplified by practical exercises using artificial data provided by MS. In respect to the practical exercises it has to be noted that DoS has no experiences in using the Software R, so the main focus should be on understanding and interpretation of output from analysis.

Strengthening the capacity of Jordan's Department of Statistics

The subjects that will be covered are:

- Why is small area statistics needed
- Definition and theory behind SAE methodology
- Identification of the informative gaps with respect to the use small area methodology – both in respect to practical and technical side of small area statistics.
- Clarification of the needs
- Basic smoothing - Production of direct estimate and their variances, Production of other indirect estimators (synthetic and composite) from designs-based perspective, Assessment – required threshold, etc.
- Discusion of indicators, data availability, and data preparation
- Outline a custom-fit design/model to be used in the next mission

3. Expected output of the activity

- Activity report;
- Common understanding of definitions and concepts;
- Best practice of basic smoothing introduced;
- Practice and experiences from artificial data obtained;
- The data situation in DoS has been revied for their potential to be used as auxiliary information for producing small statistics for expenditure
- Outline a custom-fit design/model to be used in the next mission;

4. Participants

MS Short Term Experts (STE's)

- **PhD Ms. Danila Filipponi**, Head of the Labour Register Unit, Directorate for Methodology and Statistical Process Design, The Italian National Institute of Statistics (ISTAT) (MS Component Leader). Ms. Filipponi hold a PhD in Statistics and has a long and established experience in the field of model based estimators and for using administrative data as auxiliary information. In addition Ms. Filipponi has a long and established experience using R. E-mail: dafilipp@istat.it

- **PhD Mr. Andreas Berg,** Statistician Federal Statistical Office of Germany (Destatis). Mr. Berg hold a PhD in Statistics and has solid experience in the field of Small Area Estimation (SAE). As a statistician Mr. Berg is familiar with various statistical analysis packages and numerous statistical mathematical-statistical methods. E-mail: Andreas.Berg@destatis.de

Strengthening the capacity of Jordan's Department of Statistics

DoS experts (Tentative list)
- Ms. Fatima Awamreh, Head of Household Expenditure and Income Division (BC Component Leader). E-mail: Fatmeh.Awamreh@DOS.GOV.JO

- Ms. Sana Al-Momani, Household and Population Surveys Directory. E-mail: Sana.AlMomani@DOS.GOV.JO

- **Eng. Mohammad Khalaf**, Director of Sustainable Development Unit, Department of Statistics (DoS), Jordan (RTA Counterpart) and trainer in R and basic statistics. E-mail: Mohammed.Khalaf@DOS.GOV.JO

- **Ms. Roqayah Alsanabra,** Directorate of Methodologies and Data Division, Quality Assurance Division E-mail: Roqayah.Alsanabra@DOS.GOV.JO;

- **Abrar Qudah**, E-mail: Abrar.Qudah@DOS.GOV.JO;

- **Rania Abu Dhaim**, E-mail: Rania.AbuDhaim@DOS.GOV.JO;

- **Wafaa Amer**, E-mail: Wafaa.Amer@DOS.GOV.JO;

- **Walaa AlHadidi**, Walaa.AlHadidi@DOS.GOV.JO;

- **Ms. Areej Khabour**, E-mail: areej.khabour@DOS.GOV.JO;

- **Mr. Raed Salameh**, E-mail: Raed.Salameh@DOS.GOV.JO;

- **Nusaibh Abdullah,** E-mail: Nusaibh.Abdullah@DOS.GOV.JO;

- **Tamadhor Ali**, E-mail: Tamadhor.Ali@DOS.GOV.JO;

- **Ram Zayed**, E-mail: Ram.Zayed@DOS.GOV.JO;

Twinning team
- **Dr. Charlotte Nielsen** (RTA). E-mail: cln@dst.dk
- **Ms. Zaina** Amireh (Language Assistant). E-mail: zainaamireh3@gmail.com
- **Ms. Thekra Thekra Altorah** (RTA Assistant) E-mail: thekra.twinning.rtaa@gmail.com

5. Resources

Translation and interpretation will be provide throughout the activity. Translation will be provided as sequential translation. Therefore, please keep frequent pauses when presenting and talking allowing our project translator to provide as accurate a translation as possible.

All material will provided in both English and Arabic before, under and after the Mission.

Strengthening the capacity of Jordan's Department of Statistics

The venue will the Meeting room at DoS. Flip-overs and other office material will be available. DoS participants will bring laptops where R and R studio is installed

6. Overall agenda
- Day 1: Introduction to SAE
- Day 2: Indicators and Direct estimate and their variances
- Day 3: Introduction to indirect estimators (synthetic and composite) from designs-based perspective
- Day 4: Identification of SAE needs in DoS and summing up and conclussions

7. Experiences in DoS on SAE and current status

Why the need for SAE in Jordan
In Jordan small area estimation is important in the light of the increasing demand for statistical output for small geographic areas and sub-population groups. It is also important for the purposes of monitoring Sustainable Development Indicators. Traditional sample surveys design does not contribute in obtaining direct small area estimations, instead valid statistical models can provide more accurate data for small areas and populations. This component will focus on small area estimation related to expenditure in a way so the methodology, over time, can be expanded to other statistical areas and domains within DoS.

Experiences with SAE in DoS
In the past Department of Statistics (DoS) has in particularly adressed Small Area Estimation (SAE), in relation to poverty are derived from the Household Expenditure and Income Survey (HEIS), using weights from the Population and housing census from 2015. The estimation was based on methods introduced in a former project by the Word Bank using STATA as the analytic tool. The HEIS was designed at the 12 governates in Jordan. However, the method is not appropriate for developing statistics for smaller areas, as required by policy makers. The situation has been exacerbated by the influx of Syrian refugees who are included in the Population and housing census totals, but where updated information is lacking since 2015, on their movements within the Kingdom and their current place of residence. Reliable representative samples, need to be developed for the purposes of compiling robust small area estimation. This component focuses on developing statistical methodologies for estimating small area statistics on expenditure.

Currently available administrative data sources in DoS

Inventory of administrative data for persons and households
- Civil Status and Passport Department (44 variables available)
- Social Security (6 variables available)
- Ministry of Higher Education and Science (16 variables available)
- Ministry of Education (16 variables available)
- Ministry of Health Insurance Department (15 variables available)

Strengthening the capacity of Jordan's Department of Statistics

Inventory of administrative data for businesses
- The Ministry of Industry and Trade (MIT) (16 variables)
- The Companies Control Department (CCD) (10 variables)
- The Social Security Corporation (SSC) (28 variables)

Inventory of administrative data for Dwelling and housing
- The Ministry of Local Administration (x variables)
- Greater Amman Municipality (x variables)

Metadata will be provided before the Mission

National standardized hierarchical coding for geographical location used in Jordan

In Jordan a national agreed on hierarchical coding for are used by all administrative data sources. The code consist of 15 digits to place the units to the geographical levels listed below (Map for Governorate are shown to the left):

- 3 digits – Country code (e.g. 101 for Jordan)
- 2 digits – Governorate (Altogether there as 12 Governorates in Jordan (can be clustered into Amman; South and North)
- 2 digits – District
- 1 digit – Sub-district
- 3 digits – City/Village
- 2 digits – Area
- 2 digits – Sub-Area

Figure 1: Hierarchical coding of geographic information available used in all administrative data sources in Jordan.

Basic training provided before the Mission

Basic training provided for 12 staff members in DoS by Eng. Mohammad Khalaf, Director of Sustainable Development Unit, Department of Statistics (DoS) before re-starting component 2 – all topics has been explified using R.

1. Basic elements of probability theory. Concepts of sample space, events, probability axioms, and properties of probability and conditional probability.
2. Random variables. Discrete and continuous random variables, their distribution functions, expected values, variances, covariances, and other characteristics.
3. Main probability distributions. Binomial, Poisson, Gaussian, exponential and other distributions. Their properties and possible applications.
4. Sampling distributions. Distribution of the sample mean, sample variance, and sample proportions. Central limit theorem and its implications.
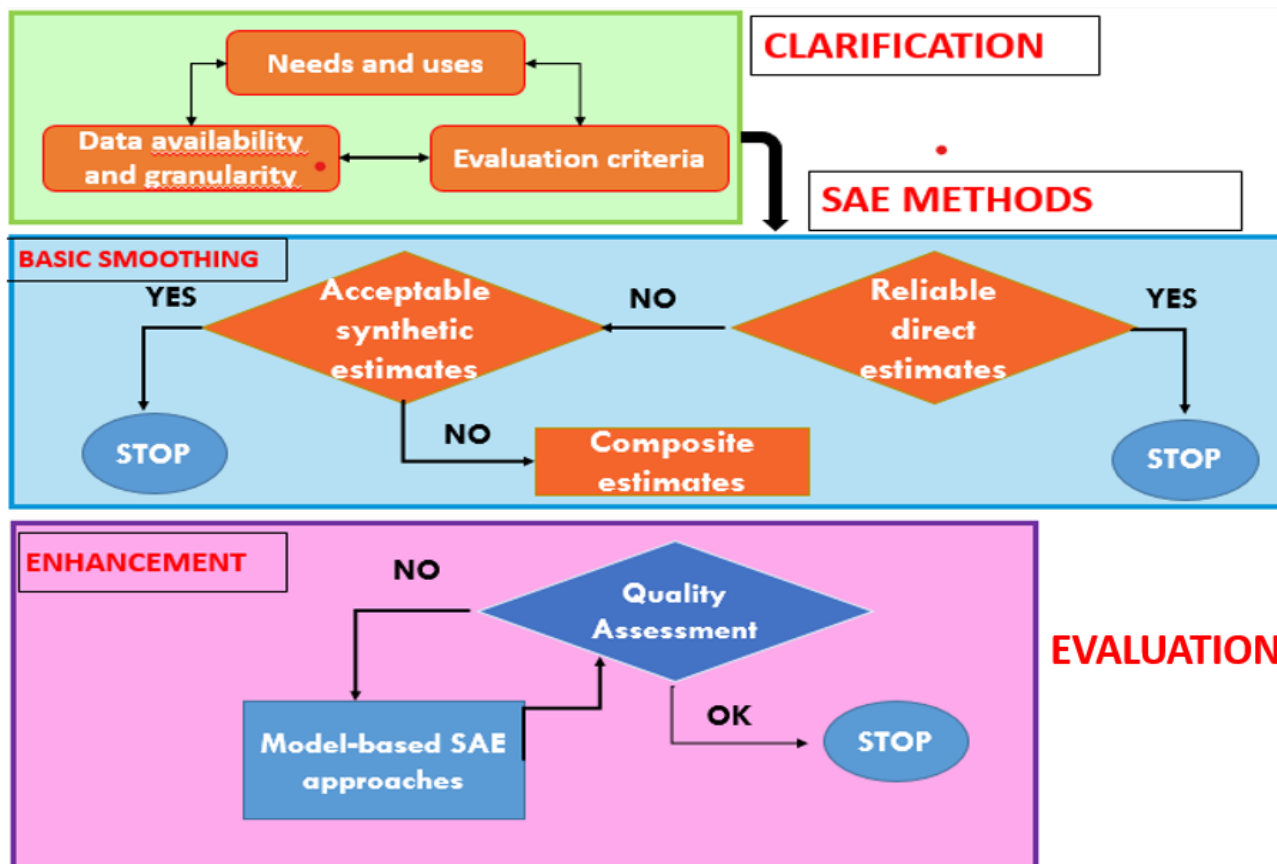
Strengthening the capacity of Jordan's Department of Statistics

5. Estimation. Point estimation and interval estimation. Methods for estimating (infinite) population parameters, including maximum likelihood estimation and method of moments. The properties of estimators such as unbiasedness, efficiency, and consistency.
6. Hypothesis testing. Formulating and testing statistical hypotheses. Common tests like t-tests and chi-square tests.
7. Confidence intervals. Constructing confidence intervals for (infinite) population parameters.
8. Descriptive statistics. Main measures of central tendency and dispersion, percentiles, frequency distributions, cross-tabulation, and graphical summaries.
9. Simple and multiple linear regression. Introduction to regression analysis, model assumptions such as linearity, independence, and homoscedasticity/heteroscedasticity, least squares estimation, weighted least squares estimation, interpretation of regression coefficients, goodness-of-fit measures, hypothesis testing in regression, and model selection techniques.
10. Mixed models. Understanding simple linear mixed models.

R training provided before the Mission

R Tutorial	What is R Programming
	Why use R Programming
	Applications of R
	Installation of R
	R Studio IDE
	Installation of R Studio
	R Advantages and Disadvantages
	R integration and Hadoop
R Basics	Syntax of R Programming
	Data types in R Programming
	Data Structures in R Programming
	Variables in R Programming
	Keywords in R Programming
	Operators in R Programming

R Statements	R IF Statement
	If-else Statement
	R Switch Statement
	R Next Statement
	R Break Statement
R Loops	R For Loop
	R Repeat Loop
	R While Loop
R Functions	R Functions
	R Built-in Functions
R Data Structures	R Vector
	R Lists
	R Arrays
	R Matrix
	R data Frame
	R Factors

Strengthening the capacity of Jordan's Department of Statistics

Process flow used for SAE introduced in DoS

Strengthening the capacity of Jordan's Department of Statistics

Annex 2: Programme for the mission

Day 1

| 9.30 -12.00 | Welcome |
| | Introduction to SAE |

| 13.00-15.00 | Practical examples |
| | Indicators, data availability and data preparation |

Day 2

| 9.30 -12.00 | Summary of day 1 |
| | Introduction to direct estimate and their variances |

| 13.30-15.00 | Direct estimate and their variances – practical examples in R with synthetic data |
| | Indicators, data availability and data preparation |

Day 3

| 9.30 -12.00 | Summary of day 2 |
| | Introduction to indirect estimators from designs-based perspective |

| 13.00-15.00 | Indirect estimators – practical examples in R with synthetic data |
| | Indicators, data availability and data preparation |

Day 4

9.30 -12.00	Summary of day 3
	Follow up from previous days
	Identification of SAE needs in DoS
	Indicators, data availability and data preparation

| 13.00-15.00 | Summing up and conclusion |

Strengthening the capacity of Jordan's Department of Statistics

Annex 3. Persons met

- Ms. Fatima Awamreh, Head of Household Expenditure and Income Division (BC Component Leader). E-mail: Fatmeh.Awamreh@DOS.GOV.JO

- Ms. Sana Al-Momani, Household and Population Surveys Directory. E-mail: Sana.AlMomani@DOS.GOV.JO

- **Eng. Mohammad Khalaf**, Director of Sustainable Development Unit, Department of Statistics (DoS), Jordan (RTA Counterpart) and trainer in R and basic statistics. E-mail: Mohammed.Khalaf@DOS.GOV.JO

- **Ms. Roqayah Alsanabra,** Directorate of Methodologies and Data Division, Quality Assurance Division E-mail: Roqayah.Alsanabra@DOS.GOV.JO;

- **Abrar Qudah**, E-mail: Abrar.Qudah@DOS.GOV.JO;

- **Rania Abu Dhaim**, E-mail: Rania.AbuDhaim@DOS.GOV.JO;

- **Wafaa Amer**, E-mail: Wafaa.Amer@DOS.GOV.JO;

- **Walaa AlHadidi**, Walaa.AlHadidi@DOS.GOV.JO;

- **Ms. Areej Khabour**, E-mail: areej.khabour@DOS.GOV.JO;

- **Mr. Raed Salameh**, E-mail: Raed.Salameh@DOS.GOV.JO;

- **Nusaibh Abdullah,** E-mail: Nusaibh.Abdullah@DOS.GOV.JO;

- **Tamadhor Ali**, E-mail: Tamadhor.Ali@DOS.GOV.JO;

- **Ram Zayed**, E-mail: Ram.Zayed@DOS.GOV.JO;

Annex 4. Course material

- Slides
- Syntetic data
- R Code