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 2.1.1
Theory and best practice of Small area estimations – Part 1
Basic smoothing and data manipulation

Mission carried out by
Dr. Andrius Čiginas, Statistics Lithuania. State Data Agency
Dr. Michele D'Alo, The Italian National Institute of Statistics (ISTAT)
Mr. Tomas Rudys, Statistics Lithuania. State Data Agency

Amman, Jordan
12-16 February 2023

Version: 2.0

دائرة الإحصاءات العامة
Department of Statistics
STATISTICS DENMARK
DESTATIS Statistisches Bundesamt
Istat
STATISTICS LITHUANIA STATE DATA AGENCY
Statistics Finland

Authors' names, addresses, e-mails

Dr. Andrius Čiginas (MS Component Leader)
Specialist at Methodology and Data Science Group
Statistics Lithuania
Gedimino Ave. 29
01500 Vilnius
Lithuania
Tel:+37065668785
Email: andrius.ciginas@stat.gov.lt

Dr. Michele D'Alo
Head of a team in charge of small areas estimation and integration of data sources
The Italian National Institute of Statistics (ISTAT)
Via C. Balbo, 15
00184 Roma
Italy
Tel: +3946733026
Email: dalo@istat.it

Mr. Tomas Rudys
Head of Methodology and Data Science Group
Statistics Lithuania. State Data Agency
Gedimino Ave. 29
01500 Vilnius
Lithuania
Tel:+37068781037
Email: tomas.rudys@stat.gov.lt

Table of contents

List of Abbreviations

- BC – Beneficiary Country
- CoP – General Population and Housing Census 2015
- DoS – Department of Statistics
- HEIS – Household Expenditure and Income Survey
- MS – Member State
- PL – Project Leader
- RTA – Resident Twinning Advisor
- STE – Short-term Expert
- SAE – Small area estimation

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". This was the first activity in Component 2; Methodology for producing Small Area Statistics titled 'Theory and best practice of Small area estimations – Part 1 Basic smoothing and data manipulation'. The actions planned for this activity were carried out as scheduled.

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

The subjects that were covered:

- Introduction to basic R usage.
- Definition and theory behind SAE methodology.
- Identification of the informative gaps with respect to the use of small area methodology.
- Basic smoothing: production of direct estimates and their variance estimates, production of the indirect estimates (synthetic and composite) from a design-based perspective, quality assessment, etc.
- Clarification of the needs.

The consultants would like to express their sincere thanks to all officials and individuals they met for the kind support and valuable information they received during their 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 EU and Statistics Lithuania – State Date Agency, and The Italian National Institute of Statistics (ISTAT).

2. Assessment and results

- Short introductory course to R was given:
    - General introduction to R (language and environment for statistical computing and graphics) and CRAN (The Comprehensive R Archive Network) was given.
    - Different possibilities to install base R were presented. Practical examples of writing scripts in console were presented.
    - Different R graphical user interfaces (GUI) were introduced. Installation of RStudio GUI was done. Live demonstration of main RStudio functionalities were done. Participants were introduced were to find information regarding working with RStudio.
    - Participants were introduced with R and GSBPM relation and where to find particular software for particular subprocesses regarding GSBPM.
    - Sources were to find information related to particular topic in R were presented (CRAN task views, Cheat sheets, etc.).

- Participants were introduced how to work with R packages and use additional functions for easier manipulation with packages.
- Related statistical tools developed with R which are not available in CRAN were presented.
- Basic R syntax were presented (concepts of R objects, scripts in external files, working directories, projects, assignment operators, vectors, mathematical functions, sequences, logical vectors, missing values, character vectors, index vectors, factors, arrays, matrices, lists, data frames, conditional sentences, loops, possibilities to write own functions, data import, export).

- Introduction to SAE and SDG indicators:
  - the importance of disaggregation for SDG indicators,
  - fundamental concepts for direct and indirect estimation,
  - auxiliary data role,
  - the process flow for producing SAE,
  - concept behind basic smoothing methods and a way to verify the results.

In this first operational mission on small area estimation methods a focus was made on:

- some types of SDG indicators, which can be defined to study the welfare of countries at different levels of disaggregation;
- the definition of small areas;
- the process to be followed to obtain estimates for unplanned domains;
- what information can be used to overcome for the lack of domain-specific information;
- the way in which extra information can be used at the estimation stage.

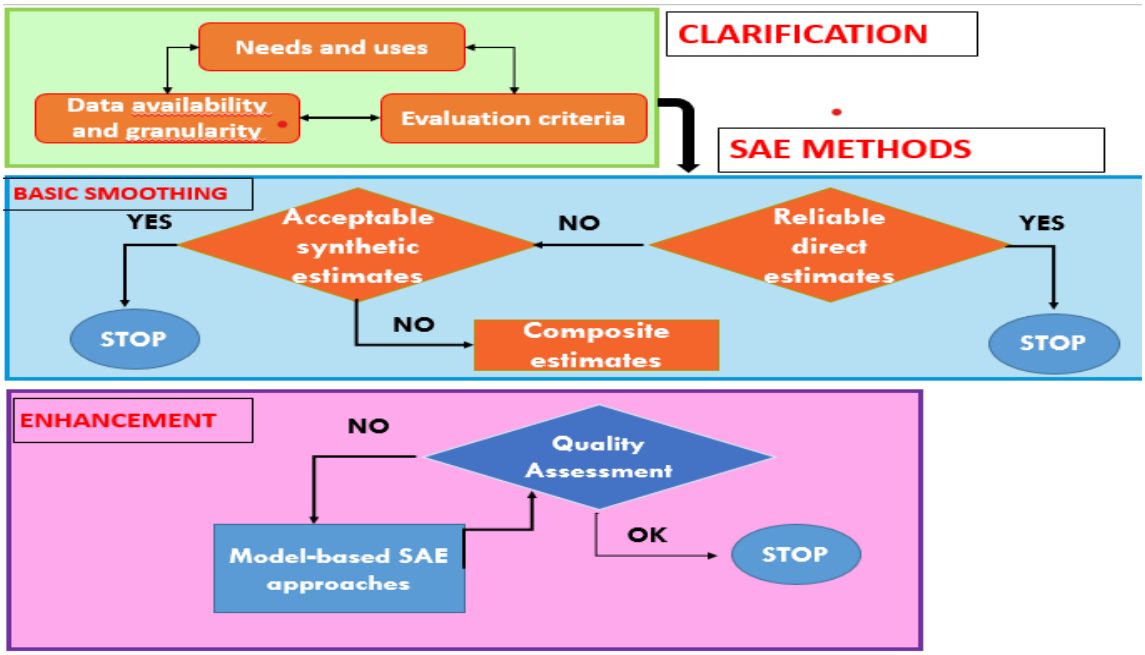
Statistics provide important inputs for evidence-based policymaking, for instance, statistics on household income and/or expenditures are commonly collected for poverty estimation and mapping. The Sustainable Development Goals (SDGs) have provided a global development framework for monitoring the social, economic, and environmental aspects of sustainable development at very disaggregated level.

Granular statistics according to some socio-economic characteristic, geographic location, and other relevant dimensions can facilitate more efficient policy because they allow to better identify specific subgroups of the population that need specific development actions. Of course, greater is the disaggregation level required bigger is the information needed to compute efficient estimates of interest.A problem of small area estimation arises when "direct estimates" based on the domain-specific data are too unreliable to be accepted in general.

A complete list of the SDG Indicators divided in 17 main dimension can be found in; https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202022%20refinement_Eng.pdf. In this project we will be focus on the indicators of the first dimension- Goal 1 and in particular on the development of Small Area Estimation (SAE) methods to be applied to some SDG indicators on poverty. The ultimate goal is to provide Jordan Department of Statistic (DoS) of the tools needed to deal with the issue of estimating parameters of interest at desired detailed level of disaggregation.

During this mission a process to be followed in order to deal small area estimation has been proposed. It can be schematized as follows:

Figure 1 – A process for computing SAE



The flowchart shows the steps of the process that invole first of all (fiirst block of the flowchart) a precise and detailed identification and prioritization of needs and uses of SAE; the assessment of data availability in terms of both survey and auxiliary data. Moreover, when developing a SAE implementation plan it is important to carefully examine the characteristics of the variables of interest for which disaggregated estimates need to be produced. The definition of the target parameter to be estimated is of vital importance. The target parameter needs to be well-supported by the available data. Indicators can have different functional forms that needs to be considered when choosing the appropriate SAE approach, and an increasing complexity of the indicators of interest simultaneously increases the granularity of the data that are needed for the estimation. Another important issue is related to the level of disaggregation for which estimates need to be computed. Sub-populations can be identified by geographic areas, but they can also represent specific socio-demographic groups. Hence, estimation domains are sub-populations of the target population that can be identified by:

- Geographic areas (governorates, districts, sub-district);
- Socio-demographic groups (sex, age, race)
- Other sub-populations (e.g. the set of firms belonging to an industry subdivision)
- A Cross classifications of the above domains

Small domain/area is estimation domain where the direct estimator does not reach a pre-specified level of precision due to 1) small sampling size or 2) absence of sampling observations.

The next step (block 2 of the flowchart) involves the computation of direct estimates even for unplanned domains and a verification of the level of the precision of these estimates. Direct estimates should always be calculated for all domains (planned and not) that are covered by the sample.

Assessment of the accuracy of direct estimates are usually performed using the Coefficient of Variation (CV) of the estimates. CV is below under a specific threshold are considered fairly accurate estimates. For instance Statistics Canada uses three categories of reliability for the Labour Force Survey: no release restriction for a CV ≤ 16.5% , added warning when 16.5% < CV≤ 33.3% and otherwise, the data is not recommended for release. Sometimes only estimates below 20% are considered reliable.

When the basic direct estimator does not allow to provide reliable estimates auxiliary data – whenever possible – can be used to improve the reliability of the estimates (decrease Variance). There are many ways of using auxiliary information to improve direct estimates' precision (e.g. Ratio estimator, post- stratified ratio estimator, Regression estimator, Calibration estimator), but also using the so called indirect estimators.

The indirect estimators may offer greater precision than the direct estimators by exploiting observed values of the target variables for a larger area, a broad-area H˜ , containing the small area d, and/or sample values which are collected for different time occasions besides the current one.The indirect estimators may be classify in synthetic and composite estimators and may be further classified according to the underlying inferential approach, i.e. design-based, model-assisted or model-based. When the indirect estimators are derived under a model-based approach you can have a model-based synthetic estimator, if a fixed model is assumed or a model based composite estimator, if a mixed model is taken into account. In the first two approaches the target parameters are considered as unknown but fixed quantities, whereas in the third one a model is formalized in order to obtain the estimator. The properties of design based and model- assisted approaches are derived on the design randomization as with direct estimator. In the third approach, the parameters of interest are considered random variables; inference is based on statistical models which describe the relationship between the variables of interest and auxiliary information. Before computing model based SAE estimators, simple indirect estimators should be calculated following the design-based approach. This can help at least useful guidance on how to continue, expecially for for practitioners who are unfamiliar with small area estimation. Simple design based synthecitc estimator can be always compute. Here some examples from Guidelines on SAE ESSnet on SAE (2012) https://ec.europa.eu/eurostat/cros/system/files/WP6-Report.pdf).

:

- The broad area ratio estimator of a small area total given the product of the within-area population size and an estimated mean of a 'broader' area to which the small area belongs. The estimator is feasible without any additional auxiliary information.
- The "simple" ratio-synthetic estimator of a small area total is given the product of the within-area auxiliary total and an estimated ratio between the target variable and the auxiliary variable, from a 'broader' area to which the small area belongs.
- The count-synthetic estimator is based on area-specific post-stratification, where the post-stratum means are taken from a 'broader' area to which the small area belongs. •

> The "combined" ratio-synthetic estimator is obtained if a "simple" ratiosynthetic estimator is applied inside each post-stratum. The auxiliary categorical information for post-stratification is combined with an additional auxiliary variable.

Synthetic indirect estimates can be more precise than the direct but can be severely biased is the implicit assumptions on which they are based are not hold. In this case, composite estimators, built as a linear combination between a direct and a synthetic estimator represents a good compromise in terms of efficiency and bias of the two components. It can computed under a design or model based approach. The main SAE estimator are usually composite type of estimator.

The aim of this mission was on the is to give a theoretical introduction of small area estimation with a focus on basic smoothing methods, then the third block of the flowchart will be a topic of future missions.

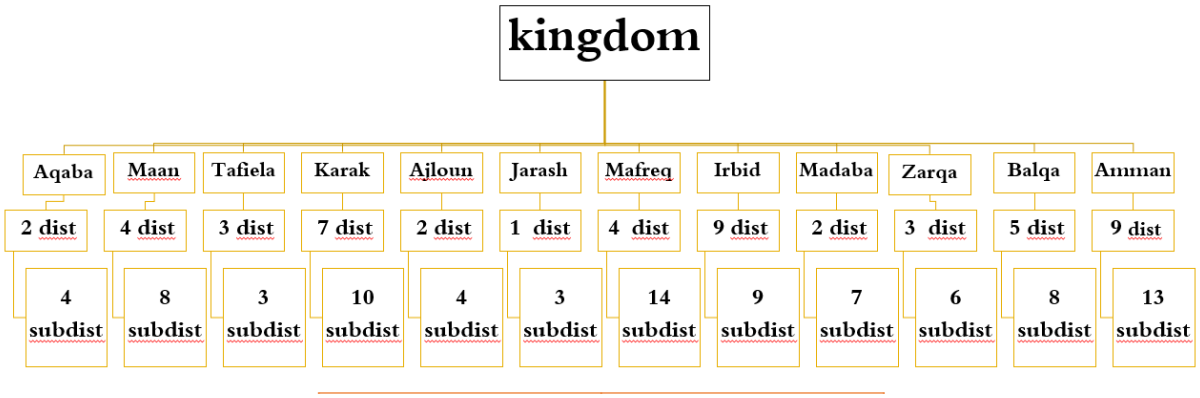
With reference to the specific needs within the Dos it is useful to underlying that the parameters of interest identified by DoS during this mission are: the **poverty rate (**for this aspect DoS should specify better the indicator they are interest in**)**; the **average annual household expenditure** and the **average annual household income**, that can be estimated by means Jordan Household Expenditure and Income Survey (HEIS). The main characteristics collected with the survey refers to the following aspects:

- Dwelling Characteristics
- Availability of durable goods and household ownership
- Social Protection
- Characteristics of individuals
- Agricultural activities
- sources of income
- Expenditures data

The HEIS survey is based on a two-stage stratified sample design. In the first stage 2,500 Primary Sampling Units (PSUs) were selected with probability proportionate to size, using as a sample frame the list of 2015 Census blocks.In the second stage 8 households (– 6 Jordanians and 2 Non-Jordanians) per PSU are drawn using systematic sampling method. Then theoretical sample has a size of 20,000 households is allocated into the 12 governorates in order to bring about accurate estimates both for each governorate and for the Kingdom as a whole. In addition, the sample cover all 51 districts and 89 subdistricts. The planned domains are identified by the 12 Governorates and by the classification of the population in the two classes of Jordanians and non-Jordanians people, the districts and the sub-district are unplanned domains.

DoS is interest on computing estimates also for the 51 districts and the 89 sub-districts that are unplanned domains in HEIS survey. The 89 sub districts are a partition of the 51 districts which constitute, in turn, a partition of the 12 governorates as shown in the figure 2..

Figure 2 Geographical Partition of the Jordanian Kingdom



It seems that direct estimates of the parameters of interest are computed from dos using the classic Horvitz-Thompson estimator. No direct estimation method that makes use of auxiliary information is currently applied for calculating estimates at the planned domain level.

No direct estimates seem to have ever been calculated for unplanned domains. It appears that the only attempt to compute estimates for these domains has been done by applying directly the ELL method proposed by the World Bank on the basis of the information collected by the 2017/2018 HEIS survey and using the 2015 Census information.

In order to follow the proposed process of the figure 1 and before described, DoS should apply the direct estimator they apply for the computing estimates at planned domains (National and governorates level) also for computing estimates of the three parameters, poverty rate; the average annual household expenditure and the average annual household income at districts and sub-districts domains.

In order to follow the proposed process of the figure 1 and before described, DoS should apply the same type of direct estimator used for the computing estimates for planned domains (National and governorates levels) also for computing estimates of the three target parameters - i.e. poverty rate; average annual household expenditure and average annual household income- at districts and sub-districts domains (unplanned domains). Then, a quality assessment in terms of CV should be performed. It is useful to remember that these domains of interest, although not planned, are all covered by the sample. Therefore for each of the 51 districts and for each of 89 sub-districts direct estimates can be computed along with their variance.

In the following table the household sample size in each Governorates are reported. The same information in terms of sampled individuals should be available and made known. Anyway, assuming that the average household size in Jordan is 4.8 people, the total number of individual sampled with the HEIS survey is about of 92,237. At the same way an approximated sample size in terms of individual can be derived for the 12 governorates.

Table 1- *Households sample size in the 12 Governorates for 2017 HEIS survey*

Governorate	Census block sample size	Household sample size	Approximated individual sample size
Amman	702	5.323	25.550
Balqa	152	1.205	5.784
Zarqa	252	1.997	9.586
Madaba	133	1.063	5.102
Irbid	314	2.478	11.894
Mafraq	145	1.113	5.342
Jerash	135	1.048	5.030
Ajloun	133	1.012	4.858
Karak	139	1.080	5.184
Tafilah	130	1.004	4.819
Maan	132	948	4.550
Aqaba	133	945	4.536
Total	**2,500**	**19.216**	**92.237**

For these planned domains the direct estimates of the target parameters and their variance should be available inside DoS and this information **should be shared within the component 2 of the project.**

Since PSU, households and individual sampled for HEIS survey cover also all 51 districts and 89 subdistricts, **direct estimates of the target parameters and their variance can and must be computed also for these unplanned domains, before applying any kind method of estimation for small areas, especially model based SAE methods.**

In the following tables household and the approximated individual sample size for 51 districts and 89 sub-districts

Table 2 - *Household and the approximated individual sample size in Jordinian Districts*

Districts	Household sample size	Approximated individual sample size
Amman Qasabah	1.159	5.563
Marka	1.277	6.130
Quaismeh	749	3.595
Al-Jamiah	1.018	4.886
Wadi Essier	522	2.506
Sahab	199	955
Jizah	126	605
Muaqqar	103	494
Naoor	170	816
Salt Qasabah	391	1.877
Shoonah Janoobiyah	116	557
Dair Alla	167	802
Ain Albasha	428	2.054
Mahes & Fuhais	103	494

Zarqa Qasabah	1.169	5.611
Russeifa	708	3.398
Hashemiyah	120	576
Madaba Qasabah	863	4.142
Deiban	200	960
Irbid Qasabah	1.040	4.992
Ramtha	319	1.531
Koorah	222	1.066
Bani Kenanah	191	917
Shoneh Shamalieh	175	840
Bani Obeid	287	1.378
Mazar Shamali	111	533
Taybeh	71	341
Wastiyah	62	298
Mafraq Qasabah	519	2.491
Badiah Shamaliyah	252	1.210
Badiah Shamaliyah Gharbiyah	326	1.565
Rwaished	16	77
Jarash Qasabah	1.048	5.030
Ajlun Qasabah	785	3.768
Kufranjah	227	1.090
Karak Qasabah	358	1.718
Mazar Janoobi	343	1.646
Qasr	100	480
Aghwar Janoobiyah	161	773
Ayy	24	115
Faqoe	64	307
Qatraneh	30	144
Tafiela Qasabah	642	3.082
Bsaira	271	1.301
Hasa	91	437
Maan Qasabah	563	2.702
Petra	131	629
Shobak	140	672
Huseiniya	114	547
Aqaba Qasabah	800	3.840
Quairah	145	696
Total	**19.216**	**92.237**

Table 3 - *Household and the approximated individual sample size in Jordinian Sub-Districts*

Districts	**Household sample size**	**Approximated individual sample size**
Amman Qasabah	1.159	5.563
Marka	1.277	6.130
Quaismeh	749	3.595
Al-Jamiah	1.018	4.886
Wadi Essier	522	2.506

Sahab	199	955
Jizah	110	528
Um Al-Rasas	16	77
Muaqqar	56	269
Rugm al-shami	47	226
Naoor	107	514
Um Elbasatien	23	110
Hosban	40	192
Salt	280	1.344
Al-Ardha	40	192
Zay	47	226
Ira & Yarqa	24	115
Shoonah Janoobiyah	116	557
Dair Alla	167	802
Ain Albasha	428	2.054
Fuhais	103	494
Zarqa	1.033	4.958
Bierain	40	192
Dhlail	72	346
Azraq	24	115
Russeifa	708	3.398
Hashemiyah	120	576
Madaba	711	3.413
Jrainah	56	269
Maeen	56	269
Faisaliyah	40	192
Deiban	80	384
Areedh	16	77
Mlaih	104	499
Irbid	1.040	4.992
Ramtha	319	1.531
Koorah	222	1.066
Bani Kenanah	191	917
Shoneh Shamalieh	175	840
Bani Obeid	287	1.378
Mazar Shamali	111	533
Taybeh	71	341
Wastiyah	62	298
Mafraq	316	1.517
Balama	103	494
Irhab	61	293
Manshiyah	39	187
Salheiah	71	341
Sabha	45	216
Um Al-Jemal	65	312
Dair Al kahf	32	154

Om-Elqottain	39	187
Badiah Shamaliyah Gharbiyah	102	490
Sama Asserhan	61	293
Hosha	62	298
Khaldiyah	101	485
Rwaished	16	77
Jarash	922	4.426
Mestabah	71	341
Borma	55	264
Ajlun	472	2.266
Sakhrah	191	917
Orjan	122	586
Kufranjah	227	1.090
Karak	358	1.718
Mazar Janoobi	288	1.382
Moab	55	264
Qasr	76	365
Mowjeb	24	115
Ghawr Assafi	104	499
Ghawr Al-mazraa	57	274
Ayy	24	115
Faqoe	64	307
Qatraneh	30	144
Tafiela	642	3.082
Bsaira	271	1.301
Hasa	91	437
Maan	269	1.291
Iel	103	494
Jafr	46	221
Mraighah	93	446
Athroh	52	250
Petra	131	629
Shobak	140	672
Huseiniya	114	547
Aqaba	764	3.667
Wadi Araba	36	173
Quairah	115	552
Diesah	30	144
Total	**19.216**	**92.237**

For the next mission DoS should should be add to these 3 table columns showing the direct estimates of the parameters of interest (poverty rate, **average annual household expenditure** and the **average annual household income**) and the corresponding variance of these estimates and the resulting CV.

Once evaluated the precision of these estimates in terms of CV, after the next mission DoS should be able to compute more efficient direct estimators that make use of the available auxiliary information – i.e. GREG and Modified GREG estimators. For calibrated estimators DoS members need extra training.

For computing these types of direct estimators it is already necessary the use of auxiliary information, that can be available at **different levels of aggregation** :

- **Area-level data:** auxiliary data are aggregated at the level of the considered estimation domain (e.g. district, sub-district, etc.) and not available at the unit level.
- **Unit-level data:** auxiliary data available for each unit in each domain. A fundamental requirement of Unit-level SAE approaches is that auxiliary variables share the same definition in the survey and additional considered data sources (e.g. census, admin register, etc.).

Once evaluated the precision in terms of CV also these estimates, whenever they were not precise enough, DoS could switch to indirect type estimates, starting from very simple design based synthetic estimators, showed them during this mission. By using a set of this estimates, composite type of estimators, given by a linear combination of direct and synthetic estimators, such as the sample-size-dependent estimator could be also implemented. Indirect estimates borrowing strength from other area and/or sources of information in order to overcome the lack of information.

An assessment of the type of auxiliary information available should be made by DoS and used to calculate the first estimates by direct and indirect methods. The recommendations of STEs are the following. If administrative data related to the income of individuals are available already, for example, from The Social Security Authority, DoS should prepare auxiliary variables from there. It is also recommended to prepare (take) the same variables from the Population and Housing Census, which were used as auxiliary data in the ELL method proposed by the World Bank.

During the mission, also basic design-based direct, synthetic and composite estimators were presented and explained in more detail:

- Examples of the estimators and smoothing techniques under no auxiliary information, and under area-level and unit-level auxiliary information were given. It was explained how the accuracy and reliability of the estimators are evaluated.
- The application of the introduced estimators is illustrated in R using artificial LFS data, where proportions like poverty rates are estimated in small areas.

The DoS representatives were active in trying to learn new methods. However, more training is needed inside and outside the Twinning project. A good approach might be to have prepared actual HEIS and auxiliary data and learn to apply the SAE methods to them.

3. Conclusions and recommendations

- The expectations of Component 2 are ambitious and require knowledge of inferential methods, modelling, and survey sampling. Capacity building is needed within and outside the Twinning project.
- Additional knowledge about data tidying and wrangling with R is needed, that is the relevant DoS staff should be trained on or outside the Twinning project to manipulate and prepare their data for SAE applications.
- The STEs should pay attention in the next missions to applications of specific R packages to calculate the common (standard) small area estimators.
- It is recommended to establish connections with Jordan and abroad universities to do research related to SAE applications in DoS.
- DoS division staff producing survey weights should participate in future meetings because preparation and calibration of the weights have an impact on the accuracy of the survey results in general including SAE applications.
- The process flow for producing SAE statistics should be followed, that is, the design-based direct, synthetic, and composite estimators, and then model-based estimators should be elaborated sequentially.
- The STEs could demonstrate the application of SAE methods on their prepared data sets to estimate indicators like Poverty rate, Average annual household expenditure, and Average annual household income actual for HEIS. On the other hand, the HEIS team should prepare raw survey data for the forthcoming mission so that they could do estimation in parallel.
- Census data contain auxiliary information which is necessary for SAE methods. Therefore, some of the Census variables should be prepared to be used in the next mission. The administrative data obtained through Component 1 should be prepared for use as soon as possible as well.

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 2.1.1:
Theory and best practice of Small area estimations – Part 1 Basic smoothing and data manipulation

Dates: 12-16 February 2023

Content

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

0. Objective and Mandatory Results for the component

Objective

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

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 poverty and expanding the methodology to other statistical areas and domains within DoS.

Small Area Statistics (SAS), particularly in relation to poverty are derived from the Household Expenditure and Income Survey (HEIS), using weights from the CoP. The HEIS was designed at governate level and 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 CoP 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 statistics. This component focuses on developing statistical methodologies for estimating small area statistics on poverty, and expanding the methodology to other statistical areas within the DoS. SAS data are a crucially important policy requirement for the purposes of monitoring Sustainable Development Indicators (SDIs).

Recently the Jordan Economic Modernization Vision 2030 was launched and "Smart Jordan" was identified as one of the eight Growth Drivers to implement the Economic Modernization Vision. The 'Smart Jordan Driver' includes seven sectors where data is one of them. This indicates the national interest to ensure constant and reliable data sources, and robust statistical systems that contribute to timely and informed policy making. It is expected that one of the measures that will be taken is to transform Jordan's Department of Statistics (DoS) into an interactive National Statistical Center (NSC).

Component 2 is sub-divided in five sub-components each with a Mandatory Results (MR) and two to four indicators of achievements associated with the sub-component. This current mission is related MR 1.1 and MR 1.3.

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*

MR from the Twinning Fiche	Indicator
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 programmes 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

1. Lesson learned from the first STE Mission in this component

In October 2022, MS experts from Lithuania and Italy carried out a planning mission where the current situation and needs from DoS was outlined in details. The Mission report is provided in Annex C.

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.

The subjects that will be covered are:

- Definition and theory behind SAE methodology
- Identification of the informative gaps with respect to the use small area methodology.
- 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.

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;

4. Participants

MS Short Term Experts (STE's)

PhD Andrius Čiginas, Specialist at Data Technology Development Group, Statistics Lithuania. State data Agency (MS Component Leader). Mr. Čiginas is an expert in sample surveys and small area estimation for social and business statistics. He is strong in methodological and practical implementations of sample surveys and has vast experience in integrating administrative sources into surveys, and developing small area estimation modelling techniques. Mr. Čiginas has professional experience of more than 17 years in social and business sample surveys of official statistics, including methodological and hands-on experience with small area estimation in Statistics on Income and Living Conditions (SILC) and Labor Force surveys (LFS), and research based on SILC and LFS data. He has extensive experiences in delivering training.

PhD Michele d'Alo, Head of a team in charge of small areas estimation and integration of data sources, The Italian National Institute of Statistics (ISTAT). Mr. d'Alo has a long and established experience in the field of Small Area Estimation (SAE). Application of SAE methods and techniques to the main household surveys. Methodological issues related to multi-source data frameworks. Use of SAE to produce estimates of SDG indicators at a very disaggregated level. Long and established experience in the field of Small Area Estimation

(SAE). Application of SAE methods and techniques to the main household surveys. Methodological issues related to multi-source data frameworks.

Mr. Thomas Rudys, Head of Data Technology Development Group – Statistics Lithuania. State Data Agency. Mr. Thomas Rudys has a long and established experience in implementation of using R in statistical office and in providing training in R. In addition Mr. Thomas Rudys has extensive knowledge on analysis of using alternative data sources in production of official statistics; Analysis off possibilities to use Big Data, web scraped data in production of official statistics; Special statistical/mathematical/IT tools for analysis and computation for big data sets; Development of new method and technologies for production of official and experimental statistics.

DoS experts

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

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

Mr. Mohammad Al-Jundi, Director of Household and Population Surveys Directory
E-mail: Mohammad.AlJundi@DOS.GOV.JO

Mr. Mohammad Tawalbeh, Director of Agricultural and Environmental Statistics Directory
E-mail: Mohammad.Tawalbeh@DOS.GOV.JO

Mr. Mohammad Abdel Razzaq, Director of Economic Surveys Directory
E-mail: Mohammad.AbdelRazzaq@DOS.GOV.JO

Ms. Ghaida Khasawneh, Director of Methodologies and Studies Directory
E-mail: ha.Khasawneh@DOS.GOV.JO

Mr. Safwat Al-Radaideh, Head of Administrative Data Department, poverty file
E-mail: Safwat.Radaideh@DOS.GOV.JO

Mr. Mohammad Omari , Administrative Data Department
E-mail: Mohammad.Omari@DOS.GOV.JO

Ms. Ahlam Al-Rosan, Director of Electronic Transformation and Information Technology
E-mail: Ahlam.AlRosan@DOS.GOV.JO

Mr. Hussam Abu Shukur, Head of Electronic Dissemination
E-mail: Hussam.AbuShukur@DOS.GOV.JO

Twinning team

Eng. Mohammad Khalaf, Director of Sustainable Development Unit, Department of Statistics (DoS), Jordan (RTA Counterpart)
Dr. Charlotte Nielsen (RTA)
Ms. Hala Abdallat (RTA Assistant)
Ms. Zaina Amireh (Language Assistant)

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.

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 R – part 1
- Day 2: Introduction to R – part 2
- Day 3: Introduction to SAE and SDGs Indicators and Direct estimate and their variances
- Day 4: Introduction to indirect estimators (synthetic and composite) from designs-based perspective
- Day 5: Identification of SAE needs in DoS and summing up and conclussions

Annex 2: Programme for the mission

R COURSE AND THEORY AND BEST PRACTICE OF SMALL AREA ESTIMATIONS – PART 1 BASIC SMOOTHING AND DATA MANIPULATION

Andrius Čiginas , Michele D'Alò, Tomas Rudys

Sunday, February 12

9.30 -12.00 Introduction and R

13.00-15.00 R

Monday, February 13

9.30 -12.00 R

13.00-15.00 R

Tuesday, February 14

9.30 -12.00 Introduction to SAE and SDGs Indicators

13.00-15.00 Direct estimate and their variances (HT, GREG, modified GREG) (Andrius)

Wednesday, February 15

9.30 -12.00 Introduction to indirect estimators (synthetic and composite) from designs-based perspective (Andrius)

13.00-15.00 Examples and application with R

Thursday, February 16

9.30 -12.00 Identification of SAE needs in DoS. Description of HEIS surveys with regards to the indicators and the domains of interest

13.00-15.00 Summing up and conclusion