List of available traineeship positions at ISTAT in 2018

Number	Title of project	Project	Period	Place
1	Methods to evaluate the quality of statistical processes using administrative data	The traineeship starts analysing an already established statistical data production process for which a quality evaluation is required, including all its sub-processes (e.g. acquisition of administrative data, data sources management, statistical output). The task is then to understand the quality requirements and to set the appropriate evaluation procedures. The main final aim is to identify the appropriate quality components of the process, and to define and interpret the appropriate measures that can be used to detect the process potential sources of error. The traineeship will require both joint work with an Istat expert and individual work (e.g. bibliographic review).	September-October-November 2018	Rome
2	Compatibility and imputation plan for item nonresponses in population official statistics data	Definition of compatibility of data sources and imputation techniques for item nonresponses, with application to population data. Steps: definition of appropriate rules to verify data compatibility; identification of subpopulations with information errors; corrections by using imputation techniques (possible use of standard software).	June-December 2018	Rome
3	Thematic incomes register	Integration of administrative and survey data to produce estimates of incomes and transfers	October-November 2018	Rome

4	Using satellite maps to produce statistics	Due to the recent development of deep learning techniques, the use of convolutive neural networks has emerged as a robust and effective framework for advanced analysis of large quantities of visual big data. In more details, these neural networks allow features classification and segmentation from images. These features can be used to generate automated and massive statistics from a large number of images. The traineeship will focus on the classification and segmentation of satellite urban and non-urban images for the production of simple automated statistics such as land cover (i.e. land percentage occupied by civil/military buildings, trees,etc), etc. Also more complex statistics could be produced in accordance with the tutor. This theme is a critical one for Istat, since at the moment there are not alternative machine learning methodologies that can be used to analyse the large amount of satellite images that Istat can use. It is a matter of fact that it is really hard to use traditional statistical methodologies to produce aggregate automated statistics from satellite images. During the traineeship the student will be introduced to deep learning models for images classification (wide residual networks, capsule networks, etc)., segmentation (u-net, fast r-cnn, etc), to computer vision techniques (open cv), to Python programming and to several machine learning (scikit-learn) and deep learning (keras, tensorflow, etc) libraries	May-July 2018, September-December 2018	Rome

5	Machine learning approach to sentiment analysis	Sentiment analysis (SA) has recently received a growing interest since “sentiments” can be used to represent the “moods” of a population, which in turn can influence political, economic and social choices. For example SA can be used to compute benefit indicators. Istat researchers are already working on SA applied to textual big data coming from several sources (e.g. social networks such as Twitter, Facebook, etc). However, the use of cutting-edge machine learning instruments for SA is still limited in Istat. Relevant literature suggest that deep learning techniques could allow reach the state-of-the-art in accurate automatic sentiment classification (85%). A main problem in the application of “machine-learning based” SA is the scarce availability of data in the Italian language. On the other side lexicon-based approaches based on the translation from English to Italian, may suffer from severe bias, since the sentiment can vary depending on the translation. For example the same word can have a negative acceptation in English (e.g. curious), a positive one in Italian (e.g. curioso). The traineeship will focus on the proposal of new deep learning methods for sentiments classification by using training sets. Moreover, it will also focus on the development of neural machine translation models to interface the sentiment model to the Italian language. Some experiments already showed very good results. During the traineeship the student will be introduced to advanced methodologies of text classification by using deep learning state-of-art models (i.e. recurrent convolutional neural networks, recurrent neural tensor networks, etc) and neural machine translation models (seq2seq attention and beam search), to pyton programming and to several machine learning (scikit-learn) and deep learning (keras, tensorflow, etc) libraries	May-July 2018, September-December 2018	Rome

6	Semantic engine for chatbot	The development of chatbot has recently received a growing interest both in public and private research institutes. In the framework of machine learning, and in particular of deep learning, really high level results have been obtained in the last recent years. Recurrent models with sequence to sequence learning have been able to produce advanced chatbot with an interaction level that equals that of the current state-of-the-art in this sector. Many applications regards chatbot for automatic help desk, recruitment, FAQs. In all these applications the chatbot is considered as a modified and advanced version of the corresponding traditional system. The development of "statistical" chatbot is receiving a growing interest for the production of aggregate data from "chatlog". The development of a interviewer chatbot could be an interesting experiment in several areas, as it could become a new and automated instrument for the production of statistics. It could became a "2.0 Questionnaire". Moreover, seq2seq models usually applied for chatbot could also be applied for the development of semantic search engines, since they are able to extract information from training sets in a end-to-end and contextual manner. Semantic search engines are a hot issue in Istat. During the traineeship the student will be introduced to learning sequence-to-sequence methodologies (seq2seq attention models), dynamic memory networks (DMN), to pyton programming, to scraping , to dataset organization for machine translation and to several machine learning (scikit-learn) and deep learning (keras, tensorflow, etc) libraries	May-July 2018, September-December 2018	Rome

7	Machine learning approach to the time series analysis	The traineeship will focus on the development of a web application and of a time series prediction back end based on neural networks, that could be included among Istat "experimental statistics". The web application could be used in several Istat officies, since it could be used to select input columns (X) and outputs (Y) to be predicted. Moreover, it could produce time series graphical representations including predictions and the metrics employed for accuracy evaluation (RMSE, MAP, MAE, etc). During the traineeship the student will be introduced to advanced deep models for time series forecasting such as the WSAEs-LSTM, a state-of-the-art model able to efficiently combine Transformed Wavelets (TW) for "data denoising"; Stacked Autoencoders (SAE) for efficient extraction of high level features; Long Short Term Memory (LSTM) for time series modelling. Another possibility would be to study how to combine these prediction models with models for sentiments (news, Twitter, etc) classification, since this could allow an increase in prediction accuracy. Moreover, the student will be introduced to pyton programming, to scraping , to time series dataset organization and to several machine learning (scikit-learn) and deep learning (keras, tensorflow, etc) libraries.	May-July 2018, September-December 2018	Rome
8	Industrialization of statistical processes: the use of the SDMX standard	Focus of the traineeship: modelling multidimensional statistical tables, defining the appropriate metadata. Building of a dissemination database from the metadata. Publication of the datasets. Aim: spread the use of the standard by industrializing the dissemination process in the framework of the "Hub of public statistics" project. Operating method: at Istat premises in Rome with direct involvement of the student in the project	September-December 2018	Rome
9	Analysis of graphs for large scale surveys	Introduction to graphs theory; representation of data on graphs; models for graphs analysis; methods and techniques for graphs positioning. Exercises on real data.	May-December 2018	Rome

10	Educational poverty in Italy	Analysis and processing of data from social surveys and interpretation of the results.Note: preliminary knowledge of the main multivariate parametric/non-parametric statistical models would be useful (choice of the best model using a dataset and interpretation of the final results). Also some knowledge about social statistics and social surveys would be useful.	May-December 2020	Rome
11	Record linkage for Population Census experimental surveys	Population Census experimental surveys have several aims. To evaluate the quality of the survey register/list coverage it is necessary to use record linkage methods between surveys and the sample selection register. The traineeship will focus on the study and application of unit record linkage techniques.	November 2018-January 2019	Rome - Eur
12	Coverage indicators for Population Census experimental surveys	Population Census experimental surveys have several aims. To evaluate the quality of the survey register/list coverage it is necessary to use capture/recapture methods of the units to estimate unknown population totals. The traineeship will focus on the study and application of these methods.	November 2018-January 2019	Rome - Eur
13	Population projections	The student will be introduced to methodologies and applied techniques to autonomously produce official demographic projections, using a variety of options. The focus will be on the evaluation of the appropriate methodology for each demographic movement component (fertility, mortality, internal and external migration), taking into account the most updated projection modelling best practices.	February-June 2019	Rome
14	Analysis of Italian forestry surfaces	By linking available sources, the main indicators (surface, typology, geographic-altimetric localization) relative to the role of forests in the Italian territory will be reconstructed	July-December 2018	Rome - Eur
15	Statistical indicators on food safety and food security	The activity will support a project which aims at classifying the essential statistical indicators to cover nation and international knowledge needs on food security and food health	July-December 2018	Rome - Eur

16	Landscape, agriculture and rurality as touristic attraction factors	The motivational aspects connected to excursions and touristic trips are important for the sector analysts and operators. Thus, it is necessary to monitor the motivational component connected to the visited territorial characteristics by adding value to the main available informative channels on these topics.	July-December 2018	Rome - Eur
17	Estimation of water withdrawal and use in productive activities	Definition of an estimation model for water withdrawal and use by economic activity (Ateco code) at national level, and by hydrographic district. The model will integrate several information, including Istat and alternative sources, possibly also non statistical ones. The traineeship will be included in the research and statistical production activities already programmed for the 2018. Thus, it will be important to work together with Istat researchers involved in the project for the main part of the traineeship.	May-October 2018	Rome - Viale Liegi
18	Water resources evaluation and their use	Definition and estimation of hydrological indicators for quantitative evaluation of natural water resources (precipitations, actual evapotranspiration, total outflow, aquifer recharge) and for the corresponding use and exploitation. The territorial scale is the hydrographic basin. The traineeship will be included in the research and statistical production activities already programmed for the 2018. Thus, it will be important to work together with Istat researchers involved in the project for the main part of the traineeship.	May-October 2018	Rome - Viale Liegi
19	Use of GIS and WebGIS technologies in production, analysis and dissemination statistical processes	The aim of the traineeship is to deepen the theoretical and practical aspects involved in the definition and development of some modules of the GISTAT (Statistical Geographical Information System), working together with Istat technicians and researchers. These applicative modules will be the base to update, publish and analyse census micro-zones and sections. A particular attention will be devoted to data linkage.	May-December 2018	Rome - Eur

20	Comparative analysis of microdata access at national and international level	The traineeship will focus on the current mode of access to microdata in Istat and at international level, with the aim of identifying new possible modes of access for Istat. The new mode will be also defined by considering users requests.	May-December 2018	Rome
21	Development of new electronic forms for microdata access	A new online request system for microdata access will be developed by converting the paper based forms in electronic based forms. The traineeship will be mainly applicative, and, after a review of the already available instruments, it will focus on a real data application.	May-December 2019	Rome
22	Evolution of the GSIM model to tackle list code management issues	The traineeship aim is to propose a possible modification of the GSIM model in an information dissemination framework. GSIM is the new international model used to describe classifications characteristics. This model, born in a statistical framework, is strictly linked to the data collection phase, and needs to be integrated with the dissemination phase requirements. The traineeship will start with the study of the GSIM model and of the dissemination peculiarities, to proceed with the development of the new proposed model.	May 2018-May 2019	Rome
23	Data mapping for dissemination	The traineeship aim is to introduce the student to data preparation to create a dissemination data warehouse. The traineeship will also focus on an application to some datasets.	May 2018-May 2019	Rome

24	Analysis of ANNCSU (National archive of urban streets house numbers) denominations and thesaurus proposal (suggested for linguistics students).	The aim of the traineeship is to create two dictionaries of the species and naming of circulation areas to be used in the ANNCSU. This would help municipalities in the conferment, management and maintenance of the street maps. The activities will focus on: - analysis of official denominations of circulation areas in the ANNCSU; - standardization and normalization of the nomenclatures; - classification of the nomenclatures in semantic categories and sub-categories; - filing of the nomenclatures for preparation of DUG and DUF dictionaries. The traineeship will focus on the analysis of the whole current archive and on the creation of a structured database to be consulted from the web ANNCSU infrastructure.	May-December 2018	Rome - Eur
25	Study of CCNL and CCNQ (collective agreements) on working time	Study of the contractual institutions of the new CCNL and CCNQ section and union agreements and examination of the application guidelines of the ARAN (Agency for the negotiating representation of public administrations) on working time and application proposals in Istat	May-July 2018	Rome
26	Telecommute in Istat. Concilation and improvement.	Qualitative survey on telecommuters and their supervisors to understand telecommute appreciation, the suggestions to improve it, and the impact on work life balance in its positive and negative aspects	September - December 2018	Rome
27	Project for an integrated communication campaign	The aim of the traineeship is to involve the student in a concrete experience of development and management of an institutional integrated communication campaign, addressed to several targets (citizens, firms, stakeholders, public administration), characterized by cross-media and synergic use of several communication channels, and developed in non-sequential times and phases. The student, working together with Istat experts, will learn how to work “sul campo” in a complete manner, with a focus on the specific Istat area (messages, community, reference networks) at national and international level.	June-December 2018	Rome

28	The replacement rate: design study for the calculation of the indicator using administrative data	The aim of the traineeship is to verify the possibility of calculating the replacement rate (ratio of the first income annuity of the old-age pension over the last annuity of work income), using administrative data sources and in particular the "Casellario Centrale delle Pensioni" dataset (Istat) and the "Banca Dati Reddituale" (MEF). The student will work on administrative microdata.	May-July 2018	Rome - Viale Liegi
29	Involvement in the research activities and preparation of the "Report on the territory"	Aim of the traineeship: study of national and international statistical data sources; data analysis and representation techniques.	September-December 2018	Rome
30	Quality indicators on different sources for tourism statistics	The sources available on the web will be analyzed using web scraping techniques, with the aim of finding the appropriate linking procedures among the different sources and among these sources and the data collected using the current official statistic survey models.	October 2018 - January 2019	Naples
31	Efficiency analysis of maritime transport companies	The aim is to study the evolution of the economic efficiency of the companies operating in this sector, which is currently in a reorganization phase due to the international dynamics.	May-September 2018	Naples
32	Outlet markets for North-Est Italian companies	Proposal of indicators for the analysis of potential markets, of the characteristics of foreign trade operators and of the exporting companies operating in the North-Est Italian regions using official Istat statistics.	June 2018 -May 2019	Venice
33	Structure and competitivity of multinational firms in Italy	Analysis of structural and economic indicators of multinational firms with the aim of evaluating the positive/negative effects on the Italian productive system	June 2018 -May 2019	Venice
34	Well-being, wealth and employment	Search, selection and processing of Istat data with the aim of defining a synthetic indicator able to measure the well-being level in the North, Centre and South of Italy, and in the provinces of the Veneto region. Istat data for the projects URBES and BES of the provinces will be the starting point.	June 2018 -May 2019	Venice
35	Statistical literacy	Design and experimentation on the field of new development modes for statistical literacy in the schools, in agreement with the national indications on kindergarten and first education cycle curricula	June 2018 -May 2019	Venice
36	Population and territory	Population dynamics of Italian regions' inhabitants by studying time series data on internal and external migration flows.	June 2018 -May 2019	Venice

37	Permanent Census of the population	Analysis of monitoring data of the data collection of the population census. Definition and computation of useful indicators to measure the response rates, also depending on the different survey modes.	June 2018 -May 2019	Venice
38	Census of public institutions	Analysis of monitoring data on the data collection of the census of public institutions. Definition and computation of useful indicators to measure institutions response rates.	June 2018 -May 2019	Venice
39	Consumer confidence	Computational Models of Tax Evasion with Heterogeneous Agents	June-September 2018	Venice
40	Social networks and segregation	Homophily and Segregation in Social Networks when Individuals are Limited Forward-Looking	September-December 2018	Venice
41	A study of public institutions	The student will support the researchers during the survey of public institutions, in the collection of data phase, to study and learn the data collection modes, the peculiarities and complexities of public institutions; the student will then analyze the results and study the public institutions sector	June-September 2018	Bologne
42	Methods of analysis to measure inequalities in well-being levels	The empirical evidence of equitable and sustainable well-being levels shows relevant territorial inequalities in Italy. The aim is to use spatial analysis methods and variability measures to derive an interpretative framework of the most significant territorial inequalities indicators.	November 2018-March 2019	Bologne
43	Data collection process of the public institutions Census	Analysis of the data collection process. Monitoring of the process. Analysis of the response rates trend.	May-September 2018	Florence
44	Data collection process of the permanent population Census	Analysis of the data collection process. Monitoring of the process. Analysis of the response rates and of the response modes trends.	Spetember-December 2018	Florence
45	Comparison of statistical indicators: the Italian regions among the European regions	Identification of the main European statistics at regional level. Analysis of some selected indicators (comparison between Tuscany and European regions).	May-July 2018	Florence
46	Analysis of administrative data on micro-enterprises	Short-term business analysis of micro-enterprises based on administrative data sources.	May-December 2018	Florence

47	Analysis of Equitable and Sustainable Well-being (BES) at territorial level	The theme can be declined in several methodological and applicative studies, that can be chosen also in connection with the student background, such as: a) methodological aspects: quality evaluation of the data sources, definition and quality evaluation of territorial indicators, methods and models for the territorial analysis of multidimensional phenomena; b) thematic and socio-economic analysis: territorial trends in unidimensional well-being components; study of the interrelation and trade-off among well-being components; evaluation of the relevance, completeness and pertinence of the framework for the measurement of the different BES domains; relation between well-being and economy (private and public). c) use of well-being measures to evaluate local policies: economic planning of local and territorial authorities (DEFR and DUP); social reporting of the mandate and of the area; multilevel governance and participatory processes; relevance and effectiveness of local welfare in the perspective of territorial well-being.	September 2018-June 2019	Ancona
48	Record linkage of samples of individual and household surveys using the administrative microdata integration system (SIM)	The aim of the traineeship is to make the student able to conduct and document a complete micro-data record linkage process (project, realization, validation) on data coming from individual and household surveys using the SIM integration base (Integrated System of Micro-data), by assigning to each sample record a unique code that allows to link survey data to data coming from one or more administrative sources. Note: it is necessary some base knowledge of the SQL language and of SAS or R programming.	October-December 2018	Rome - Eur
49	Productive structure of multinational companies	Involvement in the profiling process of some multinational companies, to define the inclusion in national accounts of the financial flows connected to their activities. A special attention will be devoted to the localization of intangible assets (Intellectual Property Products, assets) and of income flows.	September 2018-April 2019	Rome

50	The satellite account of the non-profit sector	Involvement in the analysis of funding models of the non-profit sector in Italy and in-depth analysis of NGOs structure.	September 2018-June 2019	Rome
51	Financial transactions in public finance	Development of an integrated database on the flows coming from the European Union as Structural funds and on the corresponding National co-financing amounts. In-depth analysis on the final beneficiary typology and study of the cash flows in the financial statements of involved public authorities.	October 2018-April 2019	Rome
52	Verification of the check and of the economic nature of the units eligible for S13	Involvement in the economic and behavioral analysis of specific institutional units (enterprises, public institutions, non-profit) to verify the belonging to the public administration sector	October 2018-April 2019	Rome
53	Administrative sources for environmental accounting	Development of an information base on the main environmental regulations which are relevant for the identification if administrative sources for the production of environmental accounts. A background in environmental and/or juridical subjects is advisable.	May 2018-November 2018	Rome
54	Study of the sources on eco-industries	Study on the enlargement of information sources used to compile eco-industries accounts (goods and environmental services sector) within European environmental economic accounts.	September 2018-February 2019	Rome
55	Study on envirnomental taxation	The aim of the traineeship is to deepen the knowledge on environmental taxation to improve the estimates of environmental taxes for economic activity, within the European environmental economic accounts.	May-July 2018	Rome