International Program in Survey and Data Science
Areas and Courses
November 14, 2016

Course	Description	ECTS/ Credits
	Research Questions	
Fundamentals of Survey and Data Science (SURV400)	The course introduces the student to a set of principles of survey and data science that are the basis of standard practices in these fields. The course exposes the student to key terminology and concepts of collecting and analyzing data from surveys and other data sources to gain insights and to test hypotheses about the nature of human and social behavior and interaction. It will also present a framework that will allow the student to evaluate the influence of different error sources on the quality of data.	6/3
	Data Generating Processes	
Data Collection Methods (SURV623)	This course will present research conducted to increase our understanding of how data collection decisions affect survey errors. This is not a "how–to-do-it" course on data collection, but instead reviews the literature on survey design decisions and data quality in order to sensitize you to alternative design decisions and their impact on the data obtained from surveys.	6/3
Questionnaire Design (SURV630)	The objective of this course is to introduce the scientific literature on the design, testing, and evaluation of survey questionnaires.	6/3
Introduction to Record Linkage with Big Data Applications (SURV667)	The course will address methods to combine data on given entities (people, households, firms etc.) that are stored in different data sources. By showing the strengths of these methods and by providing numerous practical examples the course will demonstrate the various benefits of record linkage. The participants will also learn about potential pitfalls record linkage projects may face.	2/1
Practical Tools for Sampling and Weighting (SURV745)	This course is a statistical methods class appropriate for second year Master's students and PhD students. The course will be a combination of hands-on applications and general review of the theory behind different approaches to sampling and weighting.	6/3
Applied Sampling (SURV625)	Methods of Survey Sampling/Applied Sampling is an applied statistical methods course, but differs from most statistics courses, though, because it is concerned almost exclusively with the design of data collection.	6/3

Course	Description	ECTS/ Credits
Causal Inference from Randomized and Observational Data (SURV722)	This course treats research designs from which causal inferences are sought. Classical experimental design will be contrasted with quasi-experiments, evaluation studies, and other observational study designs. Emphasis will be placed on how design features impact the nature of statistical estimation and inference from the designs. Issues of blocking, balancing, repeated measures, control strategies, etc., will be treated.	4/2
Sampling Theory (SURV440)	This is an introductory course in sampling theory, presenting simple random sampling, sampling for proportions, estimation of sample size, sampling with varying probabilities of selection, stratification, systematic selection, cluster sampling, double sampling, and sequential sampling.	6/3
Web Survey Methodology (SURVxxx)	The course introduces the students to the fundamentals concepts of web surveys and web survey design. The course is organized in 3 main sections which follow the way a proper web survey is organized: prefielding, fielding and post fielding.	2/1
	Data Curation and Storage	
Database Design (INST733)	Principles of user-oriented database design. Requirements analysis. Data modelling. Data integrity and security and multi-user databases. Implementing an information system using a database management system (DBMS).	6/3
Big Data Infrastructure (INST767)	Principles and techniques of data science and business intelligence. Technologies and architectures for large-scale data warehousing and scale-out data analytics platforms. Supervised and unsupervised data mining.	6/3
Principles of Data Curation (INST640)	tbd	6/3
	Data Analysis	
General Linear Models (SURVxxx)	tbd	6/3
Analysis of Complex Sample Data (SURV701)	Analysis of Complex Sample Survey Data covers the following topics: the development and handling of selection and other compensatory weights for survey data analysis; the effects of stratification and clustering on survey estimation and inference; alternative variance estimation procedures for estimated survey statistics; methods and computer software that take into account the effects of complex sample designs on survey estimation and inference; and methods for handling missing data, including weighting adjustment and imputation.	6/3
Advanced Statistical	Modeling techniques to be covered include	6/3

Course	Description	ECTS/ Credits
Modeling (SURV746)	multilevel modeling (with an application to methodological studies of interviewer effects), structural equation modeling (with an application of latent class models to methodological studies of measurement error), classification trees (with an application to prediction of response propensity), and alternative models for longitudinal data (with an application to panel survey data from the Health and Retirement Study).	
Big Data and Machine Learning I-III (SURV751)	Big Data are often used for prediction and classification tasks. Both of which can be tackled with machine learning techniques. In this course we explore how Big Data concepts, processes and methods can be used within the context of Survey Research. Throughout this course we will illustrate key concepts using specific survey research examples including tailored survey designs and nonresponse adjustments and evaluation.	2/1 each
Big Data and Machine Learning II (SURVxxx)	tbd	2/1
Big Data and Machine Learning III (SURVxxx)	tbd	2/1
Measurement error models (SURV730)	Surveys reflect the opinions or facts researchers are after only partly – the other part will be measurement error, which can seriously bias analyses of interest. To remove such biases it is essential to estimate the extent of measurement error in survey variables, which is precisely the goal of statistical measurement error modeling. In this course, we will discuss how measurement error can be defined, how its presence can be detected using specialized data collection designs and models, and how to perform error-corrected statistical analyses of substantive interest.	2/1
Bayesian Analysis (SURV798Z)	tbd	2/1
Introduction to Small Area Estimation (SURV662)	There is a growing demand to produce reliable estimates of various socio-economic and health characteristics at both national and sub-national levels. However, data availability at the sub-national (small area) level from a survey is often limited by cost and thus analysts must make the best possible use of all available information. The course will begin with a history of small-area estimation and different uses of small-area statistics in both public and private sectors. This course will provide an introduction to the main concepts and issues in small estimation and describes various approaches for estimating	6/3

Course	Description	ECTS/ Credits
	different small area parameters.	
	Data Output/Access	
Information Ethics (INST610)	tdb	6/3
Data Visualization (INST760)	tbd	6/3
Privacy and Security in a Networked World (INST611)	Evolving conceptualization of privacy and security issues in light of technological developments in the 21st century. Analysis of legal, ethical, design, and socially constructed challenges that organizations and individuals face when developing privacy and security solutions.	2/1
Data Confidentiality and Statistical Disclosure Control (SURV735)	This course will provide a detailed overview of the topic, covering all important aspects relevant for the synthetic data approach: introduction to data confidentiality, different approaches to generating synthetic datasets, possible modeling strategies and analytical validity evaluations, potential measures to quantify the remaining risk of disclosure, chances and obstacles of the idea.	4/2