

SURV - SURVEY METHODOLOGY

SURV400 Fundamentals of Survey and Data Science (3 Credits)

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.

Prerequisite: STAT100; or permission of BSOS-Joint Program in Survey Methodology department.

Restriction: Course open to SURV certificate students, SURV Advanced Special Students, and SURV undergraduate minors. Graduate students from other departments may enroll with permission from the department.

Credit Only Granted for: SURV699M or SURV400.

Formerly: SURV699M.

SURV410 Introduction to Probability Theory (3 Credits)

Probability and its properties. Random variables and distribution functions in one and several dimensions. Moments. Characteristic functions. Limit theorems.

Prerequisite: 1 course with a minimum grade of C- from (MATH240, MATH461, MATH341); and 1 course with a minimum grade of C- from (MATH340, MATH241).

Cross-listed with: STAT410.

Credit Only Granted for: STAT410 or SURV410.

SURV420 Theory and Methods of Statistics (3 Credits)

Point estimation, sufficiency, completeness, Cramer-Rao inequality, maximum likelihood. Confidence intervals for parameters of normal distribution. Hypothesis testing, most powerful tests, likelihood ratio tests. Chi-square tests, analysis of variance, regression, correlation. Nonparametric methods.

Prerequisite: 1 course with a minimum grade of C- from (SURV410, STAT410).

Cross-listed with: STAT420.

Credit Only Granted for: STAT420 or SURV420.

SURV430 Fundamentals of Questionnaire Design (3 Credits)

Introduction to the scientific literature on the design, testing and evaluation of survey questionnaires, together with hands-on application of the methods discussed in class.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

Credit Only Granted for: SURV430 or SURV630.

SURV440 Sampling Theory (3 Credits)

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.

Prerequisite: STAT401 or STAT420.

Credit Only Granted for: STAT440 or SURV440.

SURV600 Fundamentals of Survey and Data Science (3 Credits)

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.

Prerequisite: Any University of Maryland approved college-level statistics course.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department; and must have graduate student status.

Credit Only Granted for: SURV699M, SURV400 or SURV600.

SURV611 Review of Statistical Concepts (3 Credits)

Basics of probability and statistics. Students will review basic probability concepts and probability distributions, the Central Limit Theorem and hypothesis testing, and linear and logistic regression. Throughout this course, students should develop and reinforce proper statistical intuition. This includes knowing how to identify a sample and a population and applying appropriate statistical methods such as hypothesis testing, as well being able to identify different types of data and using the proper methods for each type of data. By the end of the course, students should have a strong foundation in statistics with which they can start their graduate coursework.

Credit Only Granted for: SURV699M or SURV611.

Formerly: SURV699M.

SURV612 Ethical Considerations for Data Science Research (1 Credit)

The goal of research ethics is to protect human subjects from harm when they participate in a study. In the digital age, however, what constitutes "participation" has become blurry, especially with the rise of social media platforms and other online apps and services. Furthermore, new applications of big data raise important questions about how to protect consumers from harms, and what kinds of notice and consent should be obtained. This course provides an introduction and overview of research ethics in the 21st century and evaluates the many challenges to conducting ethical research.

Credit Only Granted for: SURV699A or SURV612.

Formerly: SURV699A.

SURV613 Machine Learning for Social Science (3 Credits)

Introduction to supervised statistical learning techniques such as decision trees, random forests and boosting and discusses their potential application in the social sciences. These methods focus on predicting an outcome Y based on some learned function $f(X)$ and therefore facilitate new research perspectives in comparison with traditional regression models, which primarily focus on causation. Predictive methods also provide a valuable extension to the empirical social scientists' toolkit as new data sources become more prominent. In addition to introducing supervised learning methods, the course will include practical sessions to exemplify how to tune and evaluate prediction models using the statistical programming language R.

Recommended: Students are encouraged to work through one or more R tutorials prior or during the first weeks of the course. Some resources are listed on the syllabus.

Credit Only Granted for: SURV613 or SURV699U.

Formerly: SURV699U.

SURV615 Statistical Modeling and Machine Learning I (3 Credits)

This is the first course in a two-term sequence in applied statistical methods and machine learning that are the basis in handling complex datasets. The topics covered include: overview on the quantitative research, linear regression, analysis of variance, inference, prediction, model diagnostics and selection and resampling methods. The emphasis will be to understand and apply the methods.

Prerequisite: Must have basic R Programming skills; and must have completed a two course sequence in probability and statistics; or students who have comparable content may contact the department for permission.

Restriction: Must be in Survey Methodology (Master's) program; or permission of instructor.

SURV616 Statistical Modeling and Machine Learning II (3 Credits)

Build on material presented in Statistical Methods and Machine Learning I. Topics include: categorical data analysis, logistic regression, model selection for inference and prediction, classification using K-means and neural networks, survival analysis, principal components, and factor analysis.

Prerequisite: SURV615.

SURV617 Applications of Statistical Modeling (3 Credits)

Designed for students on both the social science and statistical tracks for the two programs in survey methodology, will provide students with exposure to applications of more advanced statistical modeling tools for both substantive and methodological investigations that are not fully covered in other MPSM or JPSM courses. Modeling techniques to be covered include 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). Discussions and examples of each modeling technique will be supplemented with methods for appropriately handling complex sample designs when fitting the models. The class will focus on practical applications and software rather than extensive theoretical discussions.

Prerequisite: SURV615 and SURV616; or permission of instructor.

Credit Only Granted for: SURV617, SURV746, or SURV699R.

Formerly: SURV699R and SURV746.

SURV621 Fundamentals of Data Collection I (3 Credits)

First semester of a two-semester sequence that provides a broad overview of the processes that generate data for use in social science research. Students will gain an understanding of different types of data and how they are created, as well as their relative strengths and weaknesses. A key distinction is drawn between data that are designed, primarily survey data, and those that are found, such as administrative records, remnants of online transactions, and social media content. The course combines lectures, supplemented with assigned readings, and practical exercises. In the first semester, the focus will be on the error that is inherent in data, specifically errors of representation and errors of measurement, whether the data are designed or found. The psychological origins of survey responses are examined as a way to understand the measurement error that is inherent in answers. The effects of the mode of data collection (e.g., mobile web versus telephone interview) on survey responses also are examined.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV622 Fundamentals of Data Collection II (3 Credits)

This is the second course in a two-semester sequence that provides a broad overview of the processes that generate data for use in social science research. Students will gain an understanding of different types of data and how they are created, as well as their relative strengths and weaknesses. A key distinction is drawn between data that are designed, primarily survey data, and those that are found, such as administrative records, remnants of online transactions, and social media content. The course combines lectures, supplemented with assigned readings, and practical exercises. The second semester builds on the discussion of survey mode during the first semester, considering the role played by interviewers in telephone and in-person surveys and their effects on the data collected. Students next are introduced to the methods for extracting and re purposing found data for social science research. Methods for the classification of text, with an emphasis on automated coding methods, are introduced and selected applications considered (e.g., coding of open-ended survey responses, classification of the sentiments expressed in social media posts). Issues in using survey data and administrative records to measure change over time (longitudinal comparisons) are explored. The term concludes with an examination of methods for evaluating the quality of both designed and found data.

Prerequisite: Permission of Instructor required; or fundamentals of Data Collection I.

SURV624 Privacy Law (1 Credit)

To acquaint the students with the origins and basic principles of privacy law mainly in Europe. Furthermore, it will contrast the European privacy foundations with the U.S. approach. At the core of this course stands the new European General Data Protection Regulation (GDPR) and its applicability to specific cases and basic principles. Moreover, the course will cover current challenges to the existing privacy paradigms by big data and big data analytics.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV625 Applied Sampling (3 Credits)

Practical aspects of sample design. Topics include: probability sampling (including simple random, systematic, stratified, clustered, multistage and two-phase sampling methods), sampling with probabilities proportional to size, area sampling, telephone sampling, ratio estimation, sampling error estimation, frame problems, nonresponse, and cost factors.

Prerequisite: Must have completed a course in statistics approved by department.

SURV626 Sampling (2 Credits)

Practical aspects of sample design. The course will cover the main techniques used in sampling practice: simple random sampling, stratification, systematic selection, cluster sampling, multistage sampling, and probability proportional to size sampling. The course will also cover sampling frames, cost models, and sampling error (variance) estimation techniques.

Prerequisite: Permission of BSOS-Joint Program in Survey Methodology department; and must have taken at least a graduate level statistics course or an undergraduate level advanced statistics course.

SURV627 Experimental Design and Causal Inference (2 Credits)

Many of the questions we are interested in as researchers and practitioners are of a causal nature. We act upon the world; how can we tell if our actions have impact? How can we decide if an intervention would get us closer to our goals? In this course, we introduce the basic concepts from causal inference and econometrics, and show what makes a valid causal claim, and what would undo it. We then demonstrate how experiments can be used to evaluate causal hypotheses, and what options are available to conduct experiments in practice. Having discussed experimental data collection, we turn to the analysis of experiments, show how this, again, is linked to the logic of causal inference, and how to work with experimental data. We discuss how to design studies so that statistical inferences are informative and reliable. Next, we cover situations in which experiments might not be possible, and show how these can be addressed through study design *ex ante* and *ex post* through analysis.

Prerequisite: Basic knowledge of data analysis. Familiarity with the R programming language and the RStudio IDE.

Recommended: Experience in the use of SAS or STATA statistical analysis software.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV630 Questionnaire Design and Evaluation (3 Credits)

The stages of questionnaire design; developmental interviewing, question writing, question evaluation, pretesting, and questionnaire ordering and formatting. Reviews of the literature on questionnaire construction, the experimental literature on question effects, and the psychological literature on information processing. Examination of the diverse challenges posed by self versus proxy reporting and special attention is paid to the relationship between mode of administration and questionnaire design.

Restriction: Must be in the Survey and Data Science Doctoral or Master's Program, or the Applied Political Analytics Master's Program; or permission of the Joint Program in Survey Methodology.

Credit Only Granted for: SURV430 and SURV630.

SURV631 Questionnaire Design (2 Credits)

This course introduces students to the stages of questionnaire development. The course reviews the scientific literature on questionnaire construction, the experimental literature on question effects, and the psychological literature on information processing. It will also discuss the diverse challenges posed by self- versus proxy-reporting and special attention is paid to the relationship between mode of administration and questionnaire design. Students will also get hands-on experience in developing their own questionnaire.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV632 Cognition, Communication and Survey Measurement (3 Credits)

Major sources of survey error-such as reporting errors and nonresponse bias-from the perspective of social and cognitive psychology and related disciplines. Topics: psychology of memory and its bearing on classical survey issues (e.g., underreporting and telescoping); models of language use and their implications for the interpretation and misinterpretation of survey questions; and studies of attitudes, attitude change, and their possible application to increasing response rates and improving the measurement of opinions. Theories and findings from the social and behavioral sciences will be explored.

SURV635 Usability Testing for Survey Research (1 Credit)

Introduces the concepts of usability and usability testing and why they are needed for survey research. The course provides a theoretical model for understanding the respondent-survey interaction and then provides practical methods for incorporating iterative user-centered design and testing into the survey development process. The course provides techniques and examples for designing, planning, conducting and analyzing usability studies on web or mobile surveys

Recommended: Students should be familiar with the basics of questionnaire design. Experience with cognitive testing is a plus, but not a requirement.

Restriction: Must be in a major within the BSOS-Joint Program in Survey Methodology department; or permission of BSOS-Joint Program in Survey Methodology department.

SURV636 Sampling II (1 Credit)

Different applications of the methods and techniques covered in the Sampling I course. This is also an applied statistics methods course concerned almost exclusively with the design of data collection rather than data analysis. The course will concentrate on sampling applications to human populations, since this poses a number of particular problems not found in sampling of other types of units. The principles of sample selection, though, can be applied to many other types of populations.

Prerequisite: SURV626 or equivalent.

Recommended: Some experience with the R statistical computing software.

Repeatable to: 1 credit.

SURV650 Economic Measurement (3 Credits)

An introduction to the field of economic measurement. Sound economic data are of critical importance to policymakers, the business community, and others. Emphasis is placed on the economic concepts that underlie key economic statistics and the translation of those concepts into operational measures. Topics addressed include business survey sampling; the creation of business survey sampling frames; the collection of data from businesses; employment and earnings statistics; price statistics; output and productivity measures; the national accounts; and the statistical uses of administrative data. Lectures and course readings assume prior exposure to the tools of economic analysis.

Prerequisite: Must have completed a course in intermediate microeconomics.

Credit Only Granted for: SURV650 or SURV699L.

Formerly: SURV699L.

SURV656 Web Survey Methodology (2 Credits)

Fundamental 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.

Prerequisite: Must have completed SURV400; or must have completed SURV623; or permission of instructor. And permission of BSOS-Joint Program in Survey Methodology department.

SURV662 An Introduction to Small Area Estimation Methods (1 Credit)

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 different small area parameters. Topics include standard design-based methods, various traditional indirect methods and the state-of-the-art small-area estimation methods that use both Bayesian and empirical best prediction methods. Monte Carlo simulation results and data analysis using available statistical software will be presented.

Prerequisite: STAT400 and STAT401.

Restriction: Permission of instructor.

SURV665 Introduction to Real World Data Management (2 Credits)

Data is omnipresent in the contemporary world coming in different shapes and sized: from survey data to found data. In order to make use of such data through analysis it is necessary first to import and clean it. This is often one of the most time consuming and difficult parts of data analysis. In this course you will learn both the conceptual steps needed in preparing data for analysis as well as the practical skills to do this. The course will cover all the essential skills needed to prepare data be it survey data, administrative data or found data.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV667 Introduction to Record Linkage with Big Data Applications (2 Credits)

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 showing how each of them are performed in practice using R, the course will demonstrate the various benefits of record linkage. Participants will also learn about potential challenges that record linkage projects may face.

Prerequisite: Basic statistical concepts; and intermediate knowledge of R.

Recommended: Familiarity with regular expressions, the R packages ggplot2 and tidyverse.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV673 Introduction to Python and SQL (1 Credit)

Basics of Python and SQL for data analysis. Students will explore real publicly-available datasets, using the data analysis tools in Python to create summaries and generate visualizations. Students will learn the basics of database management and organization, as well as learn how to code in SQL and work with PostgreSQL databases. By the end of the class, students should understand how to read in data from CSV files or from the internet and become comfortable using either SQL or Python to aggregate, summarize, describe, and visualize these datasets.

Recommended: Background knowledge in programming in Python and SQL structures.

SURV675 Modern Workflows in Data Science (2 Credits)

Large data, fast pace of production, and collaboration are hallmarks of the new data environment. In this context, researchers must have a good understanding of data workflows and they must ensure consistent and reproducible practices in order to collaborate and consistently produce insights. This course deals with some of these essential topics. We will discuss the main types of workflows in data and survey sciences and how tools such as GitHub can enhance collaboration and insure reproducibility. We will also discuss the use of reproducible documents such as Rmarkdown or Jupyter Notebooks before covering how to work with distributed data using Spark. We will finish the course by discussing the use of dashboards and how to develop such tools using R Shiny.

Prerequisite: SURV665.

Recommended: R or a good knowledge of R base and tidyverse.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

Credit Only Granted for: SURV699Y or SURV675.

Formerly: SURV699Y.

SURV699 Special Topics in Survey Methodology (1-6 Credits)

Credit according to time scheduled and organization of the course.

Organized as a lecture series on specialized advanced topics in survey methodology.

Prerequisite: Must have completed a graduate-level course in statistics or quantitative methods; and must have familiarity with survey research methods.

SURV701 Analysis of Complex Sample Data (3 Credits)

Analysis of data from complex sample designs covers: the development and handling of selection and other compensatory weights; methods for handling missing data; the effect of stratification and clustering on estimation and inference; alternative variance estimation procedures; methods for incorporating weights, stratification and clustering, and imputed values in estimation and inference procedures for complex sample survey data; and generalized design effects and variance functions. Computer software that takes account of complex sample design in estimation.

Prerequisite: SURV625.

SURV702 Analysis of Complex Survey Data (2 Credits)

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.

Prerequisite: One or more graduate courses in statistics covering techniques through OLS and logistic regression, a course in applied sampling methods.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV703 Computer-Based Content Analysis I (1 Credit)

Investigate the foundations of Natural Language Processing (NLP) as tool for analyzing natural language texts in the social sciences, thus providing an alternative to traditional ways of data generation through surveys. The course introduces general use cases for NLP, provides a guide to standard operations on text as well as their implementation in the Python-based Natural Language Toolkit (NLTK) and introduces the text mining functionalities of the WEKA Machine Learning workbench. The theory part of the course worth one credit can be supplemented by an optional project part worth another credit point.

Prerequisite: Background knowledge in programming in Python and SQL structures.

Recommended: SURV736.

SURV704 Computer-Based Content Analysis II (1 Credit)

Investigates the foundations of Natural Language Processing (NLP) as tool for analyzing natural language texts in the social sciences, thus providing an alternative to traditional ways of data generation through surveys. The course introduces general use cases for NLP, provides a guide to standard operations on text as well as their implementation in the Python-based Natural Language Toolkit (NLTK) and introduces the text mining functionalities of the WEKA Machine Learning workbench. The theory part of the course worth one credit can be supplemented by an optional project part worth another credit point.

Prerequisite: SURV703; and background knowledge in programming in Python and SQL structures.

Recommended: SURV736.

SURV706 General Linear Models (2 Credits)

The main focus of this course lies on the introduction to statistical models and estimators beyond linear regression useful to social and economic scientists. It provides an overview of generalized linear models (GLM) that encompass non-normal response distributions to model functions of the mean. GLMs thus relate the expected mean $E(Y)$ of the dependent variable to the predictor variables via a specific link function. This link function permits the expected mean to be non-linearly related to the predictor variables. Examples for GLMs are the logistic regression, regressions for ordinal data, or regression models for count data. GLMs are generally estimated by use of maximum likelihood estimation. The course thus not only introduces GLMs but starts with an introduction to the principle of maximum likelihood estimation.

Recommended: Sound understanding of Linear Regression Models, Calculus and Linear Algebra.

Restriction: Must have permission of BSOS-Joint Program in Survey Methodology.

Credit Only Granted for: SURV706 or SURV699J.

Formerly: SURV699J.

SURV720 Total Survey Error and Data Quality I (2 Credits)

Total error structure of sample survey data, reviewing current research findings on the magnitudes of different error sources, design features that affect their magnitudes, and interrelationships among the errors. Coverage, nonresponse, sampling, measurement, and postsurvey processing errors. For each error source reviewed, social science theories about its causes and statistical models estimating the error source are described. Empirical studies from the survey methodological literature are reviewed to illustrate the relative magnitudes of error in different designs. Emphasis on aspects of the survey design necessary to estimate different error sources. Relationships to show how attempts to control one error source may increase another source. Attempts to model total survey error will be presented.

Prerequisite: SURV625.

Restriction: Permission of instructor.

Credit Only Granted for: (SURV720 and SURV721) or SURV723.

Formerly: SURV723.

SURV721 Total Survey Error and Data Quality II (2 Credits)

Second part of a review of total survey error structure of sample survey data. Reviewing current research findings on the magnitudes of different error sources. Students will continue work on an independent research project which provides empirical investigation of one or more error source. An analysis paper presenting findings of the project will be submitted at the end of the course.

Prerequisite: SURV720.

Restriction: Permission of instructor.

Credit Only Granted for: (SURV720 and SURV721) or SURV723.

Formerly: SURV723.

SURV725 Item Nonresponse and Imputation (1 Credit)

Missing data are a common problem which can lead to biased results if the missingness is not taken into account at the analysis stage. Imputation is often suggested as a strategy to deal with item nonresponse allowing the analyst to use standard complete data methods after the imputation. However, several misconceptions about the aims and goals of imputation make some users skeptical about the approach. In this course we will illustrate why thinking about the missing data is important and clarify which goals a useful imputation method should try to achieve.

Prerequisite: Be comfortable with generalized linear models and basic probability theory through coursework or work experience; and familiarity with the statistical software R.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV726 Multiple Imputation (1 Credit)

This course will provide a detailed introduction to multiple imputation, a convenient strategy for dealing with (item) nonresponse in surveys. We will motivate the concept and illustrate why multiple imputation should generally be preferred over single imputation methods. The main focus of the course will be on strategies to generate (multiple) imputations and how to deal with common problems when applying the methods for large scale surveys. We will also discuss various options for assessing the quality of the imputations. All concepts will be demonstrated using software illustrations in R.

Prerequisite: Be comfortable with generalized linear models and basic probability theory through coursework or work experience; and familiarity with the statistical software R; and must have completed Surv 725 Item Nonresponse and Imputation or be familiar with the content through relevant work experience.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV727 Fundamentals of Computing and Data Display (3 Credits)

The first part of this course provides an introduction to web scraping and APIs for gathering data from the web and then discusses how to store and manage (big) data from diverse sources efficiently. The second part of the course demonstrates techniques for exploring and finding patterns in (non-standard) data, with a focus on data visualization. The course focuses on R as the primary computing environment, with excursus into SQL and Big Data processing tools.

Restriction: Must be in a major within the BSOS-Joint Program in Survey Methodology department; or permission of BSOS-Joint Program in Survey Methodology department.

Additional Information: Students without any R knowledge are encouraged to work through one or more R web tutorials prior or during the first weeks of the course.

SURV730 Measurement Error Models (1 Credit)

Measurement error in survey data can significantly distort analyses of substantive interest. Means, totals, and proportions will be off if the average answer people give is inaccurate. However, measurement error distorts not only estimates of means but can also severely bias apparent relationships, conditional probabilities, means differences, and other regression-type analyses. To remove such biases it is therefore essential to estimate the extent of measurement error in survey variables. This can be done using a gold standard or, in the absence of such a standard, modeling the error. This course introduces the latter and trains students to perform regression analyses without the influence of measurement error.

Prerequisite: SURV623 or SURV630; or have equivalent survey research experience. And must have completed a basic statistics course in regression modeling.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV735 Data Privacy and Data Confidentiality (2 Credits)

This course will provide a gentle introduction to statistical disclosure control with a focus on generating synthetic data for maintaining the confidentiality of the survey respondents. The first part of the course will introduce several traditional approaches for data protection that are widely used at statistical agencies. Some limitations of these approaches will also be discussed. The second part of the course will introduce synthetic data as a possible alternative. This part of the course will discuss different approaches to generating synthetic datasets in detail. Possible modeling strategies and analytical validity evaluations will be assessed and potential measures to quantify the remaining risk of disclosure will be presented. To provide the participants with hands on experience, all steps will be illustrated using simulated and real data examples in R.

Prerequisite: Must have familiarity with the statistical software R; and must have completed a basic statistics course in regression modeling.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV736 Introduction to Web Scraping with R (1 Credit)

Provides a condensed overview of web technologies and techniques to collect data from the web in an automated way. To this end, students will use the statistical software R. The course introduces fundamental parts of web architecture and data transmission on the web. Furthermore, students will learn how to scrape content from static and dynamic web pages and connect to APIs from popular web services. Finally, practical and ethical issues of web data collection are discussed.

Prerequisite: Students are expected to be familiar with the statistical software R.

Recommended: Knowledge about the ?tidyverse? packages, in particular, dplyr, plyr, magrittr, and stringr.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV740 Fundamentals of Inference (3 Credits)

Focuses on the fundamentals of statistical inference in the finite population setting. Overview and review fundamental ideas of making inferences about populations. Basic principles of probability sampling; focus on differences between making predictions and making inferences; explore the differences between randomized study designs and observational studies; consider model-based vs. design-based analytic approaches; review techniques designed to improve efficiency using auxiliary information; and consider non-probability sampling and related inferential techniques.

Prerequisite: SURV410 and SURV420; or (SURV615 and SURV616); or permission of Instructor required.

Restriction: Must be in a major within the BSOS-Joint Program in Survey Methodology department; or permission of BSOS-Joint Program in Survey Methodology department.

SURV742 Inference from Complex Surveys (3 Credits)

Inference from complex sample survey data covering the theoretical and empirical properties of various variance estimation strategies (e.g., Taylor series approximation, replicated methods, and bootstrap methods for complex sample designs). Incorporation of those methods into inference for complex sample survey data. Variance estimation procedures applied to descriptive estimators and to analysis of categorical data. Generalized variances and design effects presented. Methods of model-based inference for complex sample surveys examined, and results contrasted to the design-based type of inference used as the standard in the course. Real survey data illustrating the methods discussed. Students will learn the use of computer software that takes account of the sample design in estimation.

Prerequisite: SURV440.

SURV745 Practical Tools for Study Design and Inference (3 Credits)

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. Topics covered include sample size calculations using estimation targets based on relative standard error, margin of error, and power requirements. Use of mathematical programming to determine sample sizes needed to achieve estimation goals for a series of subgroups and analysis variables. Resources for designing area probability samples. Methods of sample allocation for multistage samples. Steps in weighting, including computation of base weights, non response adjustments, and uses of auxiliary data. Non response adjustment alternatives, including weighting cell adjustments, formation of cells using regression trees, and propensity score adjustments. Weighting via post stratification, raking, general regression estimation, and other types of calibration.

Prerequisite: SURV615, SURV616, and SURV625; or permission of instructor.

SURV750 Step by Step Survey Weighting (1 Credit)

Learn to calculate analysis weights for various survey designs in a real-world setting. We will cover topics on calculating base weights for single- and multistage designs, adjusting weights for unknown study eligibility and nonresponse using a few techniques, and aligning survey estimates with known population values through weight calibration. We will use specialized software for the procedures mentioned. This course will emphasize R but some examples in SAS and Stata are also discussed.

Prerequisite: SURV440, or course in sampling theory; and SURV626, or course in applied sampling.

Recommended: Some experience with variance estimation (e.g., SURV742), statistical analysis using survey data, and the R statistical computing software.

SURV751 Introduction to Big Data and Machine Learning (1 Credit)

This is an introduction to the uses and methods of working with Big data. Students explore how Big Data concepts, processes and methods can be used within the context of Survey Research.

Prerequisite: Familiarity with the statistical software R.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

Additional Information: Familiarity with model building and model selection as well as the R program is not required but could also be helpful. Students without prior knowledge in R should plan on using free online resources to make themselves familiar with the basics of this statistical programming language.

SURV752 Introduction to Data Visualization (1 Credit)

Data visualization is one of the most powerful tools to explore, understand and communicate patterns in quantitative information. At the same time, good data visualization is a surprisingly difficult task and demands three quite different skills: substantive knowledge, statistical skill, and artistic sense. The course is intended to introduce participants to a) key principles of analytic design and useful visualization techniques for the exploration and presentation of univariate and multivariate data. This course is highly applied in nature and emphasizes the practical aspects of data visualization in the social sciences. Students will learn how to evaluate data visualizations based on principles of analytic design, how to construct compelling visualizations using the free statistics software R, and how to explore and present their data with visual methods.

Prerequisite: Basic statistics understanding and bivariate linear regression.

Recommended: Experience in the use of statistical software package R.

Restriction: Permission of BSOS-Joint Program in Survey Methodology department.

SURV753 Machine Learning II (2 Credits)

Social scientists and survey researchers are confronted with an increasing number of new data sources such as apps and sensors that often result in (para)data structures that are difficult to handle with traditional modeling methods. At the same time, advances in the field of machine learning (ML) have created an array of flexible methods and tools that can be used to tackle a variety of modeling problems. Against this background, this course discusses advanced ML concepts such as cross validation, class imbalance, Boosting and Stacking as well as key approaches for facilitating model tuning and performing feature selection. In this course we also introduce additional machine learning methods including Support Vector Machines, Extra-Trees and LASSO among others. The course aims to illustrate these concepts, methods and approaches from a social science perspective. Furthermore, the course covers techniques for extracting patterns from unstructured data as well as interpreting and presenting results from machine learning algorithms. Code examples will be provided using the statistical programming language R.

Prerequisite: SURV751; or comparable knowledge or experience.

Recommended: Familiarity with the statistical programming language R.

SURV772 Survey Design Seminar (3 Credits)

Students present solutions to design issues presented to the seminar. Readings are selected from literatures not treated in other classes and practical consulting problems are addressed.

Credit Only Granted for: (SURV770 and SURV771) or SURV772.

Formerly: SURV770 and SURV771.

SURV798 Advanced Topics in Survey Methodology (3 Credits)

Individual instruction.

Repeatable to: 12 credits if content differs. Cross-listed with STAT798.

Credit Only Granted for: STAT798 or SURV798.

SURV819 Doctoral Research Seminar in Survey Methodology (1-6 Credits)

This is the first, two term seminar introducing the doctoral student to areas of integration of social science and statistical science approaches in the design, collection, and analysis of surveys.

Restriction: Permission of instructor.

SURV829 Doctoral Research Seminar in Survey Methodology (3-6 Credits)

An advanced research seminar for students preparing to do research or take doctoral comprehensive examinations.

Restriction: Permission of instructor.

Repeatable to: 6 credits if content differs.

SURV898 Pre-Candidacy Research (1-8 Credits)

SURV899 Doctoral Dissertation Research (1-8 Credits)