HOW TO CHOOSE COURSES IN THE 2020 ICPSR SUMMER PROGRAM
FIRST FOUR-WEEK SESSION

Your choice of which courses to take in a Four-Week Session of the Summer Program should be made with the following criteria in mind:

- your own substantive and methodological interests,
- your previous course work or experience in statistics, methodology, and related mathematics, and
- your subsequent teaching and research objectives.

It is important to spend a little time going beyond just a course’s title or subject area. One very helpful resource is the course syllabus, either this year’s or the one from last year. We sometimes don’t receive a course syllabus until just before the course begins; as soon as we receive a syllabus, we post it to the course description page. You can find all of last year’s and previous years’ Summer Program syllabi on our website under the tab for “Courses/Syllabi.” Please consult these syllabi to obtain detailed information about each course. In addition, the syllabi will tell you about the statistical software package(s) used in each workshop.

When selecting Summer Program courses, you should also consider suggestions from faculty members and/or colleagues at your home institution. Just be careful that these suggestions are not based on what they wish you to learn or what they want you to bring back home (in order to help them) as opposed to what you need for your future work or what you have the background to successfully master.

You will have the opportunity to discuss your course selections with a counselor during check-in and orientation on the first day of each session. During check-in, we will do our best to help you select the set of classes that meets your personal and professional needs.

One additional, important point to mention: You can change your courses during the first few days of the session. In fact, we encourage participants to “shop around” if they cannot decide between two classes. Our instructors expect some participants to sit in during the beginning of the session, so you won’t offend them! But we do recommend that you decide on your course schedule as early as possible, preferably by the third day of the session. It is important that you are in the “right” classes and able to get the most out of those classes as soon as possible.

Stated simply: We want your Summer Program training to fit YOUR needs!
What are the differences between “workshops” and “lectures”?

The workshops are the main courses in the regular four-week sessions. In general, workshops meet two hours per day, five days a week, for four weeks. This year, two workshops in the First Four-Week Session (Network Analysis I and Data Visualization) will meet for two weeks (Monday through Friday) for four hours per day providing the same number of classroom hours as the four-week workshops. The mathematics and computing lectures are supplemental courses that cover material you will need in order to be successful in the workshops.

How Many Courses Should You Take?

One of the main advantages of attending the Summer Program is the chance to take several courses on different topics. Just be careful not to overdo it. Since the material can be quite demanding on both your time and intellectual energy, it’s best to pace yourself. Don’t get burned out; at least not until the last day of the session!

Most Summer Program participants take one or two workshops, along with one or more additional lecture classes, if they need them. Almost everyone attends one of the mathematics lectures and one or more of the computing lectures, depending upon the software requirements in their workshops, research interests, and the availability of software at their home institutions

- If you decide to take two workshops per session, you may want to designate one as your “primary” course and keep up with all the work in it (i.e., attend all of the classes, participate in class, complete the assignments or exercises, etc.) throughout the entire session. You can then “audit” a second course (e.g., attend the classes but not complete some or all of the homework assignments) and still receive good exposure to the material. If it turns out that you are able to participate in both workshops at the highest level, then great!

- A few hardy souls participate in three workshops during a single four-week session. That decision is up to you, but be careful not to overextend yourself. It is better to be completely engaged in one or two workshops during the entire session than to be only partially engaged in three workshops.

- **Note:** Two First Session workshops are offered in a special two-week, four-hours per day format: Network Analysis I: Introduction is only offered June 23–July 3, and Data Visualization is only offered July 6–17. So, you can attend either one or both of these two-week workshops during the First Four-Week Session. And, you are still able to take another workshop, as well as math and computing lectures, during the entire four weeks.
We recommend that all participants attend a math lecture unless they have recently studied matrix algebra, introductory calculus, or probability distributions. The math lecture you attend should help you in your workshop(s).

- *Mathematics for Social Scientists, I* is the best complement to the *Statistics and Data Analysis I: Introduction* workshop.
- *Mathematics for Social Scientists, II* provides brief overviews of two topics: matrix algebra and calculus. The knowledge of matrix algebra is essential for all of the statistics workshops from Regression II and beyond. Calculus is useful—and some would say essential—for the MLE and Bayesian courses.
- *Mathematics for Social Scientists, III* covers probability distributions and calculus (integration). This information is useful (and, again, often considered essential) for any of our more advanced courses, such as MLE, Bayesian, and Advanced Multivariate Methods.

We also encourage you to attend one or more of the computing lecture series offered during the Four-Week Session.

- *Introduction to Computing* provides an overview to SPSS and Stata with a focus on those routines that are most frequently used in quantitative research.
- *Introduction to the R Statistical Computing Environment* covers both basic and intermediate tasks that many scholars use in R, including data manipulation, running statistical models, writing preliminary programs, and producing high quality graphics.
- *Introduction to Python* covers the basics of Python. There is no need to try to learn everything! It is more important that you feel comfortable with the software used in your ICPSR workshops, at your home institution, or for your future graduate and professional work.

I. Little or No Statistics Background, or Your First Statistics Course Was a Bust

If you have had little or no prior training in, or experience with, statistics (or if that first stats class was a bust), then this is the place to start. For beginners, don’t fret about it. We’ve all been there!

**Workshop:**

*Statistics and Data Analysis I: Introduction* provides a basic introduction to statistics, probability, and data analysis. Topics include data acquisition and management, classification, and summarization; basic probability; exploration of common distribution used in statistics; and also confidence intervals and hypothesis testing. If you are staying for the entire eight-week program, then you also should take *Statistics and Data Analysis II: The Basics of Regression* in the Second Session. (These two courses comprise an integrated sequence.)
Lectures:

Math

*Mathematics for Social Scientists I.* Some people stumble in their first statistics course because they have been away from mathematics for a long time. So, it will actually help your statistics learning to also refresh (or learn anew) the various mathematical skills covered in this class.

Computing

*Introduction to Computing,* if you need to learn the basics of either SPSS or Stata. Alternatively, if you’ve heard that everyone back in your home institution is using R, then attend the first portion of the *Introduction to the R Statistical Learning Environment* lectures.

II. Regression Analysis: The Bedrock

Regression analysis is the basis for many of the statistical techniques used in the social, behavioral, and health sciences. The Summer Program offers three levels of workshops that cover regression models; these workshops are designated I, II, and III. Each has a different intended target audience. Note that you’ll need to feel comfortable with the material in the *Statistics and Data Analysis I* workshop in order to be successful in the Regression I course.

Workshops:

*Regression Analysis I: Introduction* is best suited for those who have had a basic introduction to statistics that covered topics up to the beginning of simple bivariate regression (i.e., the usual coverage of the first-semester statistics course). The course gives a straightforward presentation of how to use and interpret multiple regression (in scalar notation). It is best suited for those who have not been exposed to the topic before or may have struggled with it in a previous course. It is also an excellent course for those who want to refresh themselves on the basic logic and application of regression analysis in order to feel comfortable with one of the main building blocks of social science research. **Note:** This course covers material that is very similar to that in the *Statistical and Data Analysis II: The Basics of Regression* workshop in the Second Session.

*Regression Analysis II: Linear Models* is one of the most popular courses in the ICPSR Summer Program, and it is the workshop that is appropriate for many graduate students. This course provides a solid and comprehensive coverage of the general linear model. It presents multiple regression in matrix form and devotes a great deal of attention to strategies for dealing with violations of the basic regression assumptions. The presentations include both the mathematical foundations and substantive applications of multiple regression. Many Summer Program participants have probably taken a similar course at their home institution (often during the first year of graduate school). Even so, a second exposure
to the subject matter is often very useful as a review. This workshop is also a “gateway” course in the sense that the material it covers is prerequisite for most of the intermediate as well as more advanced workshops in the Summer Program. **Note:** There is also a Regression II workshop in the Second Session. **TIP:** Compare both the breadth and depth of coverage in the Regression II workshop with what you have already been exposed to. Did you cover the same number of topics as this course? Did you spend as much time on each topic? And, how comfortable are you with your mastery of this material?

*Regression Analysis III: Advanced Methods* goes beyond the standard multiple regression course into new and alternative forms of analysis using graphical, nonlinear, and nonparametric techniques. This course is intended for those who feel comfortable with the general linear model and want to explore its extensions. It provides useful perspectives on many aspects of regression analysis that often do not receive much attention, although they are often encountered in everyday research (i.e., nonlinearity and outlier observations). The course takes a modern data-analytic approach and relies heavily on the use of graphical tools to facilitate more accurate and complete interpretations of regression models.

**Lectures:**

**Math**

*Mathematics for Social Scientists II* is recommended for matrix algebra for both the Regression II and Regression III courses. You won’t need it for the Regression I course, but it wouldn’t hurt to have your first exposure to it now before you take a second course in regression. The Math II lectures also give an introduction to calculus during the latter portion of the course.

*Mathematics for Social Scientists III* is recommended for distributions as well as integral calculus for the Regression III workshop.

**Computing**

*Introduction to Computing* is recommended to learn about SPSS or Stata.

*Introduction to the R Statistical Learning Environment.* You can use any of the three packages in the Regression I or Regression II workshops. You’ll need to use either Stata or R for the Regression III workshop.

**III. Regression for Categorical Outcome Variables: The Next Step**

**Workshop:**

*Maximum Likelihood Estimation I: Generalized Linear Models* (or MLE) is one of the most popular intermediate-to-advanced courses in the Summer Program. MLE is a method used to estimate the parameters in a statistical model once you have data available. The core idea
is to calculate the parameter values that maximize the likelihood function of the model. MLE is used across a range of statistical techniques but is especially important for extensions of the general linear model for categorical, ordered, and limited dependent variables—which are commonly found in social, behavioral, and health data. Thus, this course will cover such important topics as logit and probit models, both ordered and unordered dependent variables, count models, duration models, and IRT and latent class models. Note: This course covers material that is very similar to that in the *Categorical Data Analysis* workshop in the Second Session.

### Lectures:

**Math**

*M Mathematics for Social Scientists II*: Knowing matrix algebra is a must for the MLE course. Unless you’ve studied matrix algebra recently, you should attend the Math II lectures, which also give an introduction to calculus during the latter portion of the course.

*M Mathematics for Social Scientists III* covers probability distributions and integral calculus, which can be useful in the MLE course.

**Computing**

MLE uses either Stata or R.

*Introduction to Computing* for the basics of Stata.

For R, attend the *Introduction to the R Statistical Learning Environment* lectures.

### IV. Beyond Regression: Advanced Multivariate Statistical Methods

**Workshops:**

*Measurement, Scaling, and Dimensional Analysis* covers strategies for creating geometric representations of multivariate data. These methodologies are useful for data reduction, evaluating sources of variability within data, optimizing the measurement properties of a dataset, and producing graphical depictions of data. Techniques covered in this class include summated rating (or “Likert”) scales, unfolding methods, principal components, factor analysis, and multidimensional scaling. Participants taking this course should be familiar with the multiple regression model. Knowledge of matrix algebra is very useful, so consider attending the Math II lectures at the same time.

*Multivariate Statistical Methods: Advanced Topics* covers statistical techniques for dealing with multiple dependent variables in a single model. Specific techniques covered in this workshop include principal components analysis, factor analysis, canonical correlation,
cluster analysis, and MANOVA. Note that the title of this course sometimes can be a bit confusing. Many participants believe they want to learn “multivariate” techniques in order to model the effects of several independent variables on a single dependent variable. That is NOT what this course is about! Again, this workshop covers methods that are used to deal with multiple dependent variables in a single model.

*Machine Learning: Applications in Social Science Research* covers how scholars can explore “big data,” meaning massive datasets, to make better predictions on important substantive topics. Machine learning techniques can be used to uncover patterns and structure embedded in data, test and improve model specification and predictions, and perform important data reduction. Specifically, the course covers: decision trees, random forests, boosting, k-means clustering and nearest neighbors, support vector machines, kernels, neural networks, and ensemble learning. The course also deals with best practices, including error rates, cross-validation, and the use of bootstrapping methods to develop uncertainty estimates. And, it will demonstrate methods for interpreting and presenting model output.

**Lectures:**

Math

*Mathematics for Social Scientists II* for matrix algebra for the Scaling course. The Math II lectures also provide an introduction to calculus during the latter portion of the course.

*Mathematics for Social Scientists III* covers integral calculus and probability distributions, which is very useful for the Multivariate Statistics workshop.

Computing

The Scaling and Multivariate workshops both use Stata or R. The Machine Learning workshop will use R or Python. For Stata, there is *Introduction to Computing*; for R, there is *Introduction to the R Statistical Learning Environment*; for Python, there is *Introduction to Python*.

**V. Beyond Regression: Analyzing Other Types of Data**

**Workshops:**

*Time Series Analysis I: Introduction* covers regression analysis of data that have been collected over time. Because the units of analysis are sequential observations on the same entities, they cannot be regarded as a random sample. This violates some of the fundamental assumptions in regression analysis and therefore requires special methodological techniques. Participants in economics, business administration, or public policy—as well as sociology and political science—often find this an appropriate course selection.
Multilevel Models I: Introduction and Application covers regression and similar models for data that are clustered within groups (e.g., students within classes, voters in different precincts, survey respondents in different nations, etc.). Such models are known by many synonyms, including hierarchical linear models, general linear mixed models, and clustered data models. The defining feature of these models is their capacity to provide quantification and prediction of random variance due to multiple sampling dimensions (across occasions, persons, or groups, or other clusters or contextual layers such as location).

Lectures:

Math
Mathematics for Social Scientists II for matrix algebra, which is necessary for both workshops. The Math II lectures also provide an introduction to calculus during the latter portion of the course.

Mathematics for Social Scientists III covers probability distributions (as well as integral calculus), which is useful in both workshops.

Computing
The Times Series course will use Stata and R, while the MLM workshop uses R. Introduction to Computing for Stata and Introduction to the R Statistical Learning Environment for R.

VI. Beyond the Frequentist Approach: The Bayesian Paradigm

Workshop:

Bayesian Modeling for the Social Sciences I: Introduction and Application (or, simply, Bayesian Methods) is a powerful and increasingly popular methodological strategy based on likelihood methods for inference. Instead of standard frequentist analysis, Bayesian methods incorporate information from prior research into the estimation procedures as well as update the estimates as new data are observed. The course introduces the Bayesian forms of standard statistical models in linear regression and limited dependent variables regression, before dealing with measurement models, model comparison, and multilevel models. Note: The workshop assumes a very thorough understanding of multiple regression, matrix algebra, and the principles of MLE. Some calculus would also be very helpful.

Lectures:

Math
Mathematics for Social Scientists III for probability distributions and integral calculus.
Computing
The course relies on R and WinBUGS/JAGS. *Introduction to the R Statistical Learning Environment* for R. The WinBUGS/JAGS scripts will be provided in the workshop.

VII. Substantive Course: Race, Ethnicity, and Quantitative Methodology

Workshop:

*Race, Ethnicity, and Quantitative Methodology I* provides an overview of the major theories and empirical approaches to the study of intergroup attitudes. Since measurement is a key issue in the quantitative study of race and ethnicity, the workshop focuses on different methods in scaling and dimensional analyses, and their applications in the corresponding literature. The course assumes familiarity with linear regression, and so a concurrent enrollment in *Regression Analysis I: Introduction* should be sufficient. Since the focus is also on measurement theories, participants are strongly encouraged to enroll in *Measurement, Scaling, and Dimensional Analysis* or *Multivariate Statistical Methods: Advanced Topics* if they have the requisite background.

Lectures:

Math
*Mathematics for Social Scientists II* for a refresher on matrix algebra.

Computing
The workshop will use R, Stata, and SPSS, depending on the methods covered. *Introduction to Computing* for an overview of either SPSS or Stata. To learn R, please attend the *Introduction to the R Statistical Learning Environment* lectures.

VIII. Another Methodological Approach: Networks

Workshop:

*Network Analysis I: Introduction* focuses on relationships between social entities. However, the Social Network Analysis (SNA) paradigm requires a new and different set of concepts and analytic tools, beyond those provided by standard quantitative statistical models. The key idea is that the entities under investigation can interact with each other. Network analysis provides tools for representing these interactions in ways that reveal interesting characteristics of the observations. The workshop will explore basic network analysis concepts including graphs and matrices, network measures and visualization, grouping and clustering concepts, large scale networks, and change in networks. **Note 1:** Statistical inference on networks is covered in the workshop titled *Network Analysis II: Advanced Topics* in the Second Session. **Note 2:** This course focuses largely on "whole" or "complete"
networks in which sociometric analysis is required. Egocentric analysis is not a primary focus of this course but will be a topic of discussion and inclusion when appropriate within the rest of the course.

Lectures:

Math
For an introduction to matrix algebra, attend the *Mathematics for Social Scientists II* lectures.

Computing
Software packages used: UCINET, Pajek, statnet/ergm for R, and visone. To learn about R, attend the *Introduction to the R Statistical Learning Environment* lectures. The remaining packages will be provided in the workshop.

IX. Formal and Mathematical Modeling

Workshops:

*Empirical Modeling of Social Science Theory: Advanced Topics* covers how to specify, estimate, and interpret empirical models which exhibit context conditionality, temporal dependence, or spatial interdependence and endogeneity. In modern social science theory, the effects of most conditions are not constant but rather vary depending on contextual features. And, the observed outcomes and the contextual features they are embedded within both cause, and are caused in return, each other (i.e. they are endogenous). This course shows how to specify empirical models that reflect this understanding of the complex endogeneity within most social and behavioral science theories. It then shows how to estimate, interpret, and present the results of the empirical models.

*Rational Choice Theories of Politics and Society* investigates the ways that actions taken by multiple rational individual decision makers interact to generate often surprising aggregate outcomes. This workshop is an introduction to rational choice theories and their uses in the social, behavioral, or health sciences. It focuses on the logic of rational choice analysis in both explanatory and, to a lesser extent, normative contexts. The aim of the workshop is to both discuss the basic techniques of rational choice modeling and to explore the theoretical issues that motivate and limit any use of those techniques. The workshop is especially concerned with arguments surrounding how we might interpret and empirically test formal models. Note: Although the workshop does not presuppose familiarity with either game theory or the mathematics needed to solve game theoretic problems, some prior knowledge of those topics will be an advantage.

Lectures:
Math
None required but check the course syllabus for any readings that would be helpful.

Computing
The Empirical Modeling workshop will use Stata and R, so consider Introduction to Computing for an overview of Stata.

X. Special Lectures: The Hubert M. Blalock Lecture Series

The Blalock Lecture series is offered in the evenings throughout the First Session. These presentations cover a wide variety of topics in advanced quantitative methods, professional socialization, and issues surrounding diversity, equity, and inclusion. The Blalock Lectures are completely optional; no need to register in order to attend. But we hope you join us on the topics that you find interesting. Many participants find the Blalock lectures to be both informative and enjoyable! We will post information on the special Blalock lectures via email and social media during the session itself.