HOW TO CHOOSE COURSES IN THE 2018 ICPSR SUMMER PROGRAM
SECOND FOUR-WEEK SESSION

When you enroll in the ICPSR Summer Program, you choose a set of courses that you plan to attend. These selections should be based upon your own substantive and methodological interests, your previous course work in mathematics and statistics, your current capabilities, and your research objectives. Sometimes individuals make their initial course selections based upon course titles, without fully comprehending the implications of the course contents. Sometimes course selections are based upon suggestions from an advisor or fellow student. These colleagues probably have the best intentions; but, they may be identifying ICPSR courses based upon what they already know, what they wish you to learn, or what they want you to know when you return home (in order to help them). As you can imagine, this is not necessarily the course selection strategy that is best for you!

In this document, you will find an informal course-by-course discussion of what is involved to ensure the successful completion of any given class. Please take some time to review these comments. You will have an opportunity to discuss your course selections with a counselor at the in-person registration/check-in on the first day of the session. Rest assured that we will try to help you select the set of classes that best meets your personal and professional needs.

Computing

Introduction to Computing is a lecture course that will provide a basic overview of the three major statistical software packages used in the social sciences: SPSS, STATA, and SAS (they will be covered in that order). We recommend that you go to the first Computing lecture, if at all possible. On that day, the instructor will provide the schedule for the entire course. The instructor will also cover the basics of the ICPSR Summer Program computing environment (i.e., useful information for all participants). The instructors in the statistical workshops use different software packages, depending upon their own needs and interests. So you may only want to attend the Introduction to Computing sessions that deal with the packages used in your other courses. On the other hand, you are certainly welcome to attend as many of the computing lectures as you wish.

Introduction to the R Statistical Computing Environment is a separate lecture course that focuses specifically on the R statistical computing language and environment. This software is utilized in some of our more advanced workshops, including Maximum Likelihood Estimation II: Advanced Topics and Bayesian Modeling for the Social Sciences II: Advanced Topics. The R
lectures will be offered in the early evening during the first two weeks of the session. These sessions are intended for those who have had little or no prior experience with this software. But, they do cover enough material to facilitate proficiency in a software system that is often regarded as the “Lingua Franca” of modern statistics.

Mathematics

The lecture course on Matrix Algebra provides a brief overview of this important topic. Knowledge of matrix algebra is important for all of the statistical workshops from Regression II on through our most advanced courses because multivariate statistical formulas are almost always shown in matrix form. The ICPSR Summer Program does not offer a calculus course during the second four-week session (it is covered in the Mathematics for Social Scientists, II and III lectures during the first four-week session). Familiarity with calculus is advisable (and almost mandatory) for participants in Bayesian Modeling for the Social Sciences II: Advanced Topics. Please keep this in mind if you are thinking about participating in that workshop.

Participants with Little or No Background in Statistics or Mathematics

All of the courses offered during the second four-week session of the ICPSR Summer Program assume that participants have had some prior training in statistics (at least at the level of the Statistics and Data Analysis I: Introduction workshop from the first four-week session); that is, we assume you have had exposure to basic descriptive statistics, introductory probability, and statistical inference for a single sample mean (i.e., confidence intervals and hypothesis tests). If you have taken only one prior statistics course (either in the ICPSR Summer Program or elsewhere) then you should register for the Statistics and Data Analysis II: The Basics of Regression workshop. This workshop introduces regression analysis, and the level of instruction is appropriate for those who have a basic understanding of algebra, probability, association between variables, and statistical inference. Participants in Statistics and Data Analysis II are welcome to use their preferred software package, and they should attend the Introduction to Computing lectures as necessary.

Regression Analysis

The second four-week session includes two workshops that cover multiple regression analysis. Each course has a different target audience.

Statistics and Data Analysis II: The Basics of Regression provides a straightforward introduction to bivariate (or “simple”) and multiple regression. The course first covers the mathematical foundations of regression and its key assumptions. The subsequent, primary focus is on the application and interpretation of the regression model. The course also covers more advanced topics, including interactive models and logistic regression. Participants need only a basic understanding of algebra, probability, association between variables, and statistical inference. This class is most appropriate for participants who have taken one prior statistics class (either in the ICPSR Summer Program or elsewhere), those who have not been
exposed to regression in their previous coursework, or those who struggled with the topic when they were first exposed to it.

*Regression Analysis II: Linear Models* is the 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 solid and fairly 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. A course in the fundamental elements of regression models (like the *Regression Analysis I* workshop offered in the First Four-Week Session and the *Statistics and Data Analysis II: The Basics of Regression* in the Second Four-Week Session of the Program) is recommended. *Regression Analysis II* provides a very useful review of essential concepts, as well as the opportunity to apply a more comprehensive understanding of regression analysis to your own work.

**Beyond Regression: More Advanced Statistical Methods**

The *Regression Analysis II: Linear Models* workshop is not only intrinsically important for the subject matter that it covers; it is also a “gateway” course in the sense that it is a de facto prerequisite for almost all of the more advanced courses offered in the Summer Program. The statistical courses discussed below presuppose a very strong background in multiple regression and a working familiarity with basic matrix notation. This is not merely a recommendation; it should be regarded as a requirement. Those participants without a rigorous background in regression analysis will face a very steep learning curve in any of the more advanced and demanding courses covering multivariate analysis techniques.

*Longitudinal Analysis* covers statistical models for data in which there are repeated measurements of variables for the same observational units. This situation is often called “panel data analysis,” “cross-sectional time-series models,” or “mixed models.”

*Categorical Data Analysis* covers regression-like models in which one or more independent variables are used to predict a single dependent variable. Here, however, the dependent variable is either a set of two or more nominal categories (e.g. yes or no; Democrat, Independent, or Republican; passed or failed; etc.), an ordered set of categories (e.g. high, medium, low, etc.), or a count of the number of times some event occurs (e.g., the number of clients per hour in an office, etc.). This methodology is second only to regression analysis in its popularity within the ICPSR Summer Program curriculum, because such data are encountered very frequently in the social sciences. Note that this course covers material that is very similar to that in the *Maximum Likelihood Estimation I: Generalized Linear Models* workshop in the first four-week session.
Simultaneous Equation Models covers models in which several dependent variables (which may influence each other) are jointly affected by a set of independent variables. These models allow for reciprocal relationships among variables, indirect effects, and mediated influences. In the past, the analysis of simultaneous equation models was often called “path analysis” or “causal modeling.”

Structural Equation Models with Latent Variables extends the basic framework of simultaneous equation models to situations in which the variables of immediate interest are unobserved (or “latent”); instead, the researcher has empirical variables that are interpreted as indicators of the latent variables. Structural equation models are used to estimate the relationships among the unobserved variables as well as those between the unobserved variables and the observed variables. Structural equation models used to be known as “LISREL” models, after the software that was first developed to estimate their parameters.

Maximum Likelihood Estimation II: Advanced Topics assumes prior coursework not only in MLE, but also in regression and matrix algebra. This course is divided into two two-week sections. The first section covers survival/duration models and event history analysis. The second section covers analysis of cross-sectional time-series data for both continuous and discrete dependent variables.

Causal Inference for the Social Sciences covers the conditions that must exist in order for a researcher to draw valid conclusions that variation on one variable causes variation on another variable. This workshop utilizes the potential outcomes framework of causality. Topics covered include randomized experiments, observational studies, matching strategies, propensity scores, instrumental variables, difference-in-difference, and regression discontinuity.

Network Analysis II: Advanced Topics covers inferential network analysis. Specifically, how can we draw conclusions about populations of interdependent, interacting units, based upon sample observations of these units? The workshop will include discussion of models that represent complex dependent processes, exogenous covariates, latent space dimensions, longitudinal networks, and weighted networks. A first course on network analysis is a prerequisite for enrollment in this workshop.

Time Series Analysis II: Advanced Topics covers special situations involving data in which the units of analysis are sequential observations of the same entities. Topics include vector autoregression, vector error correction, state-space models, dynamic factor models, Bayesian time series models, and more. Participation in this workshop assumes successful completion of a basic course on time series analysis (equivalent to the First Session workshop on this topic offered by the ICPSR Summer Program).
Bayesian Modeling for the Social Sciences II: Advanced Topics is the most demanding course in the ICPSR Summer Program curriculum. It assumes not only a thorough background in regression, but also in maximum likelihood estimation and the basic concepts of Bayesian modeling. Participants taking this course must be proficient at both matrix algebra and calculus.

Multilevel Models II: Advanced Topics assumes some prior coursework or exposure to multilevel modeling and/or fixed and random effects in econometrics. The course also assumes a working knowledge of MLE and categorical data analysis. The workshop covers both Bayesian and likelihood approaches to multilevel modeling so some additional past exposure to Bayesian methods is helpful but not necessary. Topics include models for dealing with endogeneity (e.g. spatial and network models), generalized linear mixed models (e.g. logit, event history, and categorical), and models for complex data structures (e.g. cross-classified and multi-member models).

Formal and Mathematical Modeling
The Game Theory II: Advanced Topics workshop covers specialized topics in the study of interactions among competing actors with strategic motives. The main focus of the workshop will be dynamic games of incomplete information, including signaling games and models of bargaining and conflict with asymmetric information. Particular attention will be paid to topics and questions attracting active interest in current research in political science and economics. This workshop assumes mastery of introductory game theory (i.e., the material covered in Game Theory I during the first four-week session of the ICPSR Summer Program).

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.

The Hubert M. Blalock Lecture Series
The Blalock Lecture series is offered throughout the four-week session. These presentations cover a wide variety of topics in advanced quantitative methods (e.g., Data Mining; Statistical Graphics, and so on), race and ethnicity (e.g., previous presentations have included Latinos and the Changing of America and The Riddle of Black Conservatism), and professional socialization (e.g., Writing Grant Proposals, Academic Publishing in Books and Journals, and Teaching Statistics). The Blalock Series is completely optional. Participants should attend the
presentations, as their schedules allow, on the topics they find interesting. Note, however, that the Blalock series strives to include sessions on subjects that are of great interest to the social science research community and the presenters are excellent public speakers. For these reasons, participant reactions to the Blalock Lecture series have always been extremely positive. So, we recommend that you make them part of your own ICPSR Summer Program experience!

**How Many Courses Should You Take?**

Summer Program participants are often tempted to elect many more courses than they can safely navigate. While this might seem to be an attractive approach, it might not be the most rewarding or useful. It is physically possible to attend Summer Program classes from 9:00 a.m. until 9:00 p.m. every day. But, those who attempt such a schedule usually experience intellectual (and emotional) burn-out. ICPSR Summer Program courses are just too crammed with material, too demanding, and too time-consuming for that strategy to be successful on a regular basis.

The general guideline is that most Summer Program participants take two workshops, along with one or more additional lecture classes. With respect to the lectures, almost everyone takes one of the mathematics courses and attends at least a week or so of the computing lectures (depending upon the software requirements in their workshops). The choice of workshops, of course, depends upon your own methodological and substantive interests. Some participants take one workshop per four-week session. That is also perfectly reasonable, and would comprise what some people would consider a “full course load.”

If you decide to participate in two workshops per session, you may want to designate one as your “primary” course and keep up with all the work (i.e., attend all of the sessions, participate in class, complete homework exercises, etc.) in it throughout the entire session. You could then audit the other “secondary” course (e.g., attend the classes but not complete some or all of the homework assignments) and still receive good exposure to the material.

A few hardy souls participate in three workshops during a single four-week session. While we understand their motivation and interpret their plans as a compliment to the quality of Summer Program courses, we still caution you about choosing to follow this route. As stated earlier, three workshops would require an enormous amount of work, even if you only audit two of them. Most participants who try to do this stop attending one or more of the workshops partway through the session. And, we often find that the net results actually are less satisfactory than would have been the case if the participant had elected a smaller number of workshops at the outset.

We want you to select the combination of Summer Program courses that is most relevant and useful for your interests and professional objectives. If you are undecided about exactly which
workshops and lectures to take, you will have opportunities to talk with counselors who can advise you about your choices. We also encourage participants to “shop around” during the first day or two of the session; if you really cannot decide between two classes, then attend each one on consecutive days, and use that to guide your choice. Our instructors expect participants to do this, so you won’t offend them! We do recommend that you decide on your course schedule as early as possible—certainly no later than the third day of the session.

We hope that you find these comments useful when you are electing or amending your Summer Program course schedule. Please do not hesitate to ask or consult or email with the staff during this process. We thank you for your participation in the ICPSR Summer Program and sincerely hope that your experience with us will prove to be an experience that is uniquely positive from all perspectives—academic, professional, and social.