HOW TO CHOOSE COURSES IN THE 2015 ICPSR SUMMER PROGRAM
(SECOND FOUR-WEEK SESSION)

When you enroll in the ICPSR Summer Program through the online registration system 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. He 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.

Along with the other statistical software packages, we also offer a separate lecture course, Introduction to the R Statistical Computing Environment. This software is utilized in some of our more advanced workshops, including Advanced Topics in Maximum Likelihood Estimation, Advanced Bayesian Models for the Social Sciences, and Scaling and Dimensional Analysis. 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). This is unfortunate because familiarity with calculus is advisable (and almost mandatory) for participants in Advanced Bayesian Models for the Social Sciences. 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 Introduction to Statistics and Data Analysis I 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 Introduction to Statistics and Data Analysis II workshop. You also should be sure to attend the first week of the Introduction to Computing lectures in order to learn about SPSS software. Familiarity with SPSS is required in order to complete the homework in the introductory statistics workshop.

Regression Analysis

The second four-week session includes two workshops that cover multiple regression analysis. Each course has a different target audience. Introduction to Statistics and Data Analysis II provides a straightforward introduction to bivariate (or “simple”) and multiple regression. The focus is on application and interpretation of the regression model. 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. Note that Introduction to Statistics and Data Analysis II is very similar in content and presentation to the first-session workshop, Regression I: Introduction.

Regression II: Linear Models is the single most popular course in the ICPSR Summer Program and it is the workshop that is most 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. Many Summer Program participants have probably taken a similar course at their home institution (often during the first year of graduate school). Even if that is the case, a second exposure to the subject matter is often very useful as a review. Almost everyone who takes this workshop finds it to be a wonderful experience, both enlightening and immediately applicable to their own work.
Beyond Regression: More Advanced Statistical Methods

The Regression 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 for Generalized Linear Models workshop in the first four-week session.

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 summed 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, but not absolutely required.

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.

Advanced Topics in Maximum Likelihood Estimation 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.
Applied Multilevel Models 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).

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: 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: 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 either the four-week or five-day workshops on this topic offered by the ICPSR Summer Program).

Advanced Bayesian Models for the Social Sciences 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.

Formal and Mathematical Modeling

The Advanced Game Theory workshop covers specialized topics in the study of interactions among competing actors with strategic motives. Particular attention will be devoted to dynamic games of incomplete information, among other topics. This workshop assumes mastery of the basic game theory material (i.e., that covered in the introductory workshop on this topic during the first four-week session of the ICPSR Summer Program).

Complex Systems Models in the Social Sciences covers models that consist of a number of interacting agents, wherein each agent’s behavior is governed by a small set of simple rules. The interaction of the agents can produce complex “emergent” structures and dynamic (or adaptive) behaviors of individuals and groups. This course will involve both classroom lectures and hands-on experience in the computer lab. It will introduce “bottom-up approaches” to computer modeling and compare them to more traditional mathematical (analytical) approaches and also to “top-down” computer models (e.g., typical macro-economic models). This is a more recent, innovative, and somewhat unorthodox approach to modeling in the social sciences.
Empirical Modeling for Theory Evaluation 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.

Complex Systems Models in the Social Sciences covers models that consist of a number of interacting agents, wherein each agent’s behavior is governed by a small set of simple rules. The interaction of the agents can produce complex “emergent” structures and dynamic (or adaptive) behaviors of individuals and groups. This course will involve both classroom lectures and hands-on experience in the computer lab. It will introduce “bottom-up approaches” to computer modeling and compare them to more traditional mathematical (analytical) approaches and also to “top-down” computer models (e.g., typical macro-economic models). This is a more recent, innovative, and somewhat unorthodox approach to modeling in the social sciences.

Substantive Workshop
There is one substantive workshop offered during the second four-week session. Longitudinal Analysis of Historical Demographic Data is a competitive-enrollment class. Participation in this workshop is limited to only those specific individuals who have been admitted by the workshop administrators. Note that this course meets from 9:00 a.m. through 5:00 p.m. for the entire four-week session. Thus, participants are effectively precluded from taking any other Summer Program classes (apart from the evening Blalock Lectures).

The Hubert M. Blalock Lecture Series
This optional series of lectures and roundtables is usually presented in the evening. The Blalock sessions cover a variety of topics on relatively advanced quantitative methods, professional socialization, and sociopolitical diversity. Topics to be covered in this session include missing data, publishing in professional journals, teaching undergraduate research methods, election forecasting, and a number of other subjects.

How Many Courses?
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 one or two of the workshops, along with one or more additional lecture classes. With respect to the lectures, many participants take the matrix algebra course and attend at least a week of the computing lectures (depending upon the software requirements in their workshops). The choice of workshops depends upon your own
methodological and substantive interests. Some participants take one workshop per four-week session. That is perfectly reasonable, and comprises what many people would consider a “full course load.”

Many other participants decide to participate in two workshops per session. That, too, is a reasonable course load. If you do this, however, you may want to designate one workshop 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” workshop (e.g., attend the classes but not complete some or all of the homework assignments) and still receive good exposure to the material. This is probably the route that most Summer Program participants follow.

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.

In any case, 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.