HOW TO CHOOSE COURSES IN THE 2020 ICPSR SUMMER PROGRAM
SECOND 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. 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.

- We recommend that all participants attend the matrix algebra lectures unless they have recently studied the subject.
  - Introductory/Review Lectures on Matrix Algebra provides a background in matrix algebra for participants in ICPSR S.P. workshops. Nine one-hour lectures will be devoted to matrices and linear algebra. The lecture series is most suitable as a review for those who have been exposed to this material previously, but it is also intended to serve as a brief and limited introduction. Note that knowledge of matrix algebra is essential for all of the statistics workshops from Regression II and beyond.

- 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 every software package! 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. Only a Little Statistical Background

All of the courses offered in the Second Session 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 Session. You should have mastered the basics of descriptive statistics, introductory probability, and statistical inference for a single sample mean (i.e., confidence intervals and hypothesis tests).

Workshop:

Statistics and Data Analysis II: The Basics of Regression 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) OR
- not been exposed to regression in their previous coursework OR
- struggled with the topic when they were first exposed to it.

Lectures:

Math

Introductory/Review Lectures on Matrix Algebra. Although you will not use matrix algebra in this workshop, you will need to be familiar with it if you take more advanced statistics courses. Why not sit through it with a great teacher while you have the opportunity?

Computing

Introduction to Computing, if you need to learn the basics of either SPSS or Stata. Alternatively, if you have 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. In the Second Session, we offer two levels of workshops that cover linear regression models: *Statistics and Data Analysis II: The Basics of Regression* (see above) and *Regression Analysis II: Linear Models* (see below).

**Workshop:**

*Regression Analysis II: Linear Models*, offered in both sessions, is one of the most popular courses in the 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. **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?

**Lectures:**

Math

*Introductory/Review Lectures on Matrix Algebra* is recommended for anyone who has not recently studied matrix algebra since you will do much of the work in Regression II in matrix format.

Computing

*Introduction to Computing*, if you need to learn the basics of either SPSS or Stata. Alternatively, if you have 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.

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

**Workshop:**

*Categorical Data Analysis* is often the next course that scholars take after learning standard linear regression. The reason for this is because the course covers regression models in which the outcome variable is not continuous, but rather discrete, categorical, or limited-dependent, and much behavioral, social, and health data is of this type. Maximum likelihood estimation is the key technique for estimating categorical or limited-dependent outcome data models. The course will deal with such important topics as logit and probit models and their extensions. **Note:** This course
covers material that is very similar to that in the *Maximum Likelihood Estimation I: Generalized Linear Models* workshop in the First Session.

Lectures:

Math

*Introductory/Review Lectures on Matrix Algebra* is a must. Unless you’ve studied matrix algebra recently, you should attend the *Matrix Algebra* lectures.

Computing

*Categorical Data Analysis* uses Stata, but will also use R for those who want to use it. *Introduction to Computing* for the basics of Stata.

Alternatively, if you have 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.

IV. Beyond Regression: Data Over Time

Workshops:

*Time Series Analysis II: Introduction* covers advanced topics in time series, i.e. 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, count time series, and more. This is a second course in time series and thus participants should have successfully mastered the material in an introductory time series course, such as the *Time Series I* workshop in the First Session.

*Longitudinal Analysis* is the study of short series of observations (or repeated measurements) of the same observation units or respondents (and in which there are many units). These models are also known as panel analysis (of a cross-section of a time series) or repeated measures or growth curves (polynomials in time) or multilevel time analysis (when the levels are actually a sequence of measurements on the same respondents or units). Longitudinal analysis is used for panel surveys, experiments and quasi-experiments in health and biomedicine, repeat measurements in education and psychology, and the evaluation of prevention and treatment programs. Participants should have thorough understanding of linear regression or analysis of variance.

*Maximum Likelihood Estimation II: Advanced Topics* is a second course in the use of MLE as the main method of statistical estimation in order to analyze data that vary both over time and across units. As such, participants should have mastered both linear regression at the level of our Regression II workshops as well as the material covered in the First Session workshop *Maximum Likelihood Estimation I: Generalized Linear Models*. 


The first two-week section of the course covers methods for making inferences with repeated observations data, focusing mostly on the theory and estimation of models for panel and time-series cross-section data. Topics covered include fixed effects, random effects, dynamic panel models, random coefficient models, models for spatial dependence, and models for qualitative dependent variables.

The second two-week section of the course covers models for survival data, also called event history analysis. Survival data record the length of time until some event occurs, for example, the termination of a cabinet government or the time until an unemployment spell ends. Time-to-event occurrence is an important feature of these kinds of data, and a wide variety of survival methods, including some non-traditional models for categorical data, will be discussed.

Lectures:

Math
Introductory/Review Lectures on Matrix Algebra if you need a review of this material

Computing
The MLE II workshop will use both Stata and R. The Time Series II workshop will use Stata. Introduction to Computing for the basics of Stata. For R, attend the Introduction to the R Statistical Learning Environment lectures.

The Longitudinal workshop will use SAS. If you need to learn the basics of SAS, attend the Introduction to Computing lectures. Some examples will employ the lme/nlme libraries in the S family of packages (S/S-Plus/R). Stata has comparable capabilities for certain problems.

V. Beyond Regression: Advanced Multivariate Statistical Methods

Workshops:

Multilevel Models II: Advanced Topics extends standard multilevel models (i.e. models for data that are clustered within groups) to more complex situations such as nonlinear and nonhierarchical mixed effects models. 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). Since the workshop covers both likelihood and Bayesian approaches to multilevel modeling, participants will need to have mastered the material in:

- an introductory course on multilevel models (such as Multilevel Models I in the First Session)
- a basic maximum likelihood course or equivalent categorical data analysis course (such as the Maximum Likelihood Estimation I workshop in the First Session or the Categorical Data Analysis workshop in the Second Session).
In addition, exposure to Bayesian methods is helpful, although not required. **Note:** Multilevel models are also known by other names such as hierarchical linear models, general linear mixed models, and clustered data models.

*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”). In this situation, 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.

*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, and focuses on such research designs as randomized experiments and observational studies. Topics covered include randomized experiments, observational studies, matching strategies, propensity scores, instrumental variables, difference-in-difference, and regression discontinuity.

**Lectures:**

**Math**
- *Introductory/Review Lectures on Matrix Algebra* for a quick review of this material.

**Computing**
Causal Inference will use R. The other workshops will use both Stata and R. Simultaneous Equations will use SAS, STATA, and R. Structural Equations will use MPlus. Attend the lectures on *Introduction to Computing* for the basics of Stata and SAS. To learn about R, attend the *Introduction to the R Statistical Learning Environment* lectures.

**VI. Beyond the Frequentist Approach: The Bayesian Paradigm**

**Workshop:**

*Bayesian Modeling for the Social Sciences II: Advanced Topics* treats the theoretical and applied principles of Bayesian statistics at a second level of understanding. The bulk of the course focuses on estimating and interpreting Bayesian models from an applied perspective. Participants are introduced
to the Bayesian forms of the standard statistical models taught in linear regression and MLE courses (i.e., linear, logit/probit, Poisson, etc.). Additional topics include measurement models, model comparison, and an in-depth treatment of multilevel modeling. **Prerequisites:** This is a very demanding course that assumes not only a thorough background in regression, but also in maximum likelihood estimation and the basic concepts of Bayesian modeling. Knowledge of matrix algebra and some calculus is necessary.

**Lectures:**

**Math**

*Introductory/Review Lectures on Matrix Algebra* to brush up quickly on this material

**Computing**

The course relies on R and WinBUGS/JAGS. *Introduction to the R Statistical Learning Environment* for R. The WinBUGS, JAGS, and (briefly) Stan scripts will be provided in the workshop.

**VII. Substantive Course: Race, Ethnicity, and Quantitative Methodology, a Second Course**

**Workshop:**

*Race, Ethnicity, and Quantitative Methodology II* provides an overview of theories and methodological approaches common in research involving race and ethnicity, as well as other social identities. Participants are encouraged to identify an original research question to work on during the course. The course will discuss both qualitative and quantitative tools in the existing literature, and course participants will assess these tools to determine which is most appropriate for their project. The second part of the course will be a conference style activity where participants will present their research project/poster for feedback, which they will incorporate in their project.

Topics to be covered include: intergroup attitudes, racial and ethnic prejudice and measurement, the intersectional approach, as well as publicly available datasets in ICPSR database and elsewhere. **Note:** The course assumes statistical knowledge of linear regression. Participants unfamiliar with linear regression are encouraged to enroll in *Regression Analysis II: Linear Models*.

**Lectures:**

**Computing**

To learn about SPSS, SAS, or Stata, attend the *Introduction to Computing* lectures. To learn R, please attend the *Introduction to the R Statistical Learning Environment* lectures.

**VIII. Another Methodological Approach: Statistical Inference in Networks**

**Workshop:**
Network Analysis II: Advanced Topics covers inferential network analysis. Inferential methods of network analysis are used to test hypotheses about the generation and evolution of a network, to derive measures of uncertainty for network indices, and to find probabilistic models that accurately describe the overall features of a network. Topics include exponential random graph models, latent variable models for networks, longitudinal networks, and causal inference with networks. Mastery of the material covered in a first course in network analysis (such as the Network Analysis I in the First Session) is a prerequisite for this workshop. In addition, an understanding of maximum likelihood estimation for categorical outcome variables (such as provided by the MLE I workshop in the First Session) is necessary preparation.

Lectures:

Math
For an introduction to, or review of, matrix algebra, attend Introductory/Review of Matrix Algebra.

Computing
To learn about R, attend the Introduction to the R Statistical Learning Environment lectures.

IX. Formal Modeling

Workshops:

Game Theory introduces students to noncooperative game theory and its application in the social sciences. Topics include dominance, Nash equilibrium, subgame perfect Nash equilibrium, perfect Bayesian equilibrium, coordination games, bargaining games, signaling games, and mechanism design. The course is assessed on the basis of homework assignments and two exams, and includes guidance about both how to develop one’s own models and how to write research articles utilizing a formal model.

Social Choice Theory provides a thorough introduction to social choice theory, which is a sibling field to game theory. Social choice provides an elegant way of analyzing how groups of people interact to produce collective decisions. The mechanism by which a group of individuals with different preferences create a decision is through the use of a preference aggregation rule. The course will examine the properties of different types of rules, and then characterize the rules that yield desirable group outcomes. Prerequisites: There are no formal prerequisites for this workshop, but prior experience with a proof-based math course will be helpful. Participants should be open to learning and utilizing mathematical notation and proof techniques.

Lectures:

Math
No formal mathematics course is required for the Game Theory or the Social Choice Theory workshops, but participants should be open to learning and utilizing mathematical notation and proof techniques.

Computing
No specific software is used for the Game Theory or the Social Choice Theory workshops, although use of LaTeX to do the assignments is encouraged. See the Introduction to the LaTeX Text Processing System during the early evening of the first day of the session. To learn R, please attend the Introduction to the R Statistical Learning Environment lectures.

X. Special Lectures: The Hubert M. Blalock Lecture Series
The Blalock Lecture series is offered in the evenings throughout the Second 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.