Standard Data Analysis Series

ST 511. METHODS OF DATA ANALYSIS **(4 Credits)**
Graphical, parametric and nonparametric methods for comparing two samples; one-way and two-way analysis of variance; simple linear regression. Lec/lab.

Recommended: ST 351

ST 512. METHODS OF DATA ANALYSIS **(4 Credits)**
Multiple linear regression, including model checking, dummy variables, using regression to fit analysis of variance models, analysis of covariance, variable selection methods. Lec/lab.

Prerequisites: ST 511 with C or better
Recommended: ST 351

ST 513. METHODS OF DATA ANALYSIS **(4 Credits)**
Principles of experimental design; randomized block and factorial designs; repeated measures; categorical data analysis, including comparison of proportions, tests of homogeneity and independence in cross-classified frequency tables, Mantel-Haenszel test, logistic regression, log-linear regression. Introduction to multivariate statistics. Lec/lab.

Prerequisites: ST 512 with C or better
Recommended: ST 351

Recommended quantitative sequence for students interested in advanced statistical skills including Bayesian statistics.

We advise quantitatively-minded students with prior statistics classes, or any students who are planning a Bayesian approach to statistics in their dissertations to take a five quarter series culminating in ST559, rather than taking the ST 511 / 512 / 513 sequence.

ST 521. INTRODUCTION TO MATHEMATICAL STATISTICS **(4 Credits)**
Probability, random variables, expectation, discrete and continuous distributions, multivariate distributions.

Recommended: MTH 253

ST 522. INTRODUCTION TO MATHEMATICAL STATISTICS **(4 Credits)**
Sampling distributions, Central Limit Theorem, estimation, confidence intervals, properties of estimators, and hypothesis testing.

Prerequisites: ST 521 with C or better
Recommended: MTH 253

ST 559. BAYESIAN STATISTICS **(3 Credits)**
Bayesian statistics for data analysis. Characterizations of probability; comparative (Bayesian versus frequentist) inference; prior, posterior and predictive distributions; hierarchical modeling. Computational methods include Markov Chain Monte Carlo for posterior simulation.

Recommended: ST 562

ST 561. THEORY OF STATISTICS **(4 Credits)**
Distributions of functions of random variables, joint and conditional distributions, sampling distributions, convergence concepts, order statistics. Lec/rec.

Recommended: ST 422 or ST 522

ST 562. THEORY OF STATISTICS **(4 Credits)**
Sufficiency, exponential families, location and scale families; point estimation: maximum likelihood, Bayes, and unbiased estimators; asymptotic distributions of maximum likelihood estimators; Taylor series approximations.

Prerequisites: ST 561 with C or better
Recommended: ST 422 or ST 522

Quantitative methods in fisheries, wildlife, and natural resources

Many students sampling data from animal populations or working with assessment, decision making and other types of data analysis in the natural resources have found these applied courses to be of substantial aid above and beyond the introductory statistics sequences (ST 511/512/513 or ST 521/522).

FW 524. INTRODUCTION TO FISHERIES ASSESSMENT **(3 Credits)**
Fisheries management strategies rely on models that predict a population's responses to exploitation. This course introduces approaches commonly used to assess and

evaluate the dynamics and status of a population. Provides an overview of the terminology, data requirements, underlying rationale, assumptions, limitations and uncertainty associated with stock assessments.

Recommended: College algebra, introductory statistics and, if unfamiliar with data collection and analysis methods in fisheries, FW 454/554

FW 544. QUANTITATIVE DECISION ANALYSIS FOR FISH AND WILDLIFE MANAGEMENT **(4 Credits)**
Decision analysis allows decision makers to examine the expected effects of different strategies before implementation; incorporate multiple objectives and values of stakeholders; determine the relative influence of various sources of uncertainty; and estimate the value of collecting additional data. Quantitatively oriented graduate students in natural resources are provided with an in-depth overview of decision analysis and adaptive management, emphasizing animal population management. Lec/lab.

Recommended: ST 511 and ST 512 or equivalent, basic background in animal population dynamics and management.

FW 661. ANALYSIS OF ANIMAL POPULATIONS (5 Credits)
Quantitative methods for estimating parameters (abundance, survival, population stability) of animal populations. Emphasis is on vertebrate animals and statistical methods of hypothesis testing, parameter estimation, and inference testing. Offered odd-numbered years.

Recommended: ST 511 and ST 512 or equivalent

FES 524. NATURAL RESOURCES DATA ANALYSIS **(4 Credits)**
Hands-on experience in applied statistical modeling and data analysis for natural resources. Emphasis is on understanding of statistical models and the application and actual implementation of statistical analysis techniques, use of statistical software for analyses (e.g., R), and interpretation of findings. Students analyze data from their own research for final projects.

Prerequisites: ST 511 with B or better and ST 512 [B]

Consistently offered special topics in Fisheries and Wildlife FW599

1. **Scientific data management**, offered by Jim Powers in Winter quarter
2. **Introduction to data management and R software**, offered by Jim Peterson in Fall quarter

Population and ecological modeling

Several different departments on campus offer courses useful to graduate students interested in constructing or applying population and community models.

BEE 529. BIOSYS MODELING TECHNIQUES **(3 Credits)**
Development of mathematical models of biological and ecological systems; linear and nonlinear systems analysis; stochastic modeling and random processes; model solution and analysis techniques.

MTH 527. INTRODUCTION TO MATHEMATICAL BIOLOGY **(3 Credits)**
Modeling and mathematical analysis of biological processes using first principles at scales ranging from the molecular to the population level. Deterministic models are studied in both discrete and continuous time and analyzed using linearization principles, linear and nonlinear stability techniques, phase plane methods, and methods from partial differential equations. Results obtained from mathematical analysis will be qualitatively interpreted and applied to the biological process under investigation. All courses used to satisfy MTH prerequisites must be completed with a C or better.

Recommended: (MTH 256 or MTH 256H) and MTH 341

MTH 528. STOCHASTIC ELEMENTS IN MATHEMATICAL BIOLOGY **(3 Credits)**
An introduction to stochastic modeling of biological processes. The stochastic models covered may include Markov processes in both continuous and discrete time, urn models, branching processes, and coalescent processes. The biological applications may include genetic drift, population dynamics, genealogy, demography, and epidemiology. Mathematical results will be qualitatively interpreted and applied to the biological process under investigation. All courses used to satisfy MTH prerequisites must be completed with a C or better.

Recommended: MTH 341 and (MTH 361 or MTH 463 or MTH 563)

BEE 538. ECOLOGICAL SYSTEMS ANALYSIS (4 Credits)

An introduction to sustainability with a focus on case studies that are relevant to biological and ecological engineers. An introduction to tools that perform technical feasibility analysis, economic viability analysis, environmental risk assessment, resource sustainability assessment and life cycle assessment

(LCA). Course will consist of theory and case studies highlighting the use of LCA methods to assess sustainability.

Recommended: ENGR 391 or ENGR391H

VMB 631. MATHEMATICAL MODELING OF BIOLOGICAL SYSTEMS (3 Credits)
The use of mathematical modeling in biological sciences is studied. A variety of modeling techniques are covered including implementing the methods computationally.

Time series, spatial statistics, and multivariate statistics

These courses may be useful to students addressing questions about complex variation, including variation in time or space, or in high-dimensional datasets such as those common in systematics, community ecology, genomics and other fields.

ST 557. APPLIED MULTIVARIATE ANALYSIS **(3 Credits)**
Multivariate data structures, linear combinations; principal components, factor and latent structure analysis, canonical correlations, discriminant analysis; cluster analysis, multidimensional scaling. Not offered every year.

Recommended: (ST 412 or ST 512) and (MTH 252 or MTH 245)

ST 565. TIME SERIES (3 Credits)
Analysis of serially correlated data in both time and frequency domains. Autocorrelation and partial autocorrelation functions, autoregressive integrated moving average models, model building, forecasting; filtering, smoothing, spectral analysis, frequency response studies, Offered winter term in even years.

Recommended: (ST 412 or ST 512) and (ST 422 or ST 522)

ST 566. TIME SERIES ANALYTICS **(3 Credits)**
Focuses on statistical and analytical tools for analyzing data that are observed sequentially over time. Specific topics can vary term to term, and could include methods for exploratory time series analysis, linear time series models (ARMA, ARIMA), forecasting, spectral analysis and state-space models. The focus will be on applied problems, though some mathematical statistics is necessary for a solid

understanding of the statistical issues. This course is designed for students in Data Analytics MS and Certificate programs.

Prerequisites: ST 516 with C or better and ST 517 [C] and ST 518 [C]

ST 567. SPATIAL STATISTICS (3 Credits)
The analysis of spatial data. Graphical tools for exploring spatial data, geostatistics, variogram estimation, kriging, areal models, hierarchical spatial models, and spatio-temporal modelling. Offered winter term in odd years.

Recommended: (ST 412 or ST 512) and (ST 422 or ST 522)

BOT 570. COMMUNITY STRUCTURE AND ANALYSIS **(4 Credits)**
Quantitative methods for the analysis of biotic communities, including community concepts, estimation of community composition parameters, theoretical aspects of multivariate methods of analyzing species-importance data, and overview of multivariate tools; hands-on computer analysis of data sets. Lec/lab.

Equivalent to: BI 570
Recommended: BI 370 and (ST 412 or ST 512) and calculus

Quantitative and statistical methods for social and health science

These courses cover a variety of general statistical techniques with particular attention to applications and examples in sociology, psychology, epidemiology, biomedicine and other human or health related subdisciplines.

FES 523. QUANTITATIVE ANALYSIS IN SOCIAL SCIENCE **(4 Credits)**
Application and interpretation of statistical approaches to human dimensions of natural resources, recreation, and other social sciences. Emphasis is on an applied approach focusing on understanding data, selecting appropriate statistics for theoretical and managerial problems, using statistical software for analyses, and interpreting findings.

Recommended: (FES 522 or FOR 522) and ST 511

PSY 560. ADVANCED SOCIAL RESEARCH METHODS (4 Credits)
Advanced experimental research methods in the social sciences. Issues in psychological construct operationalization, experimental design, data collection, analysis, and report writing will be emphasized.

Recommended: PSY 301 and PSY 360

SOC 518. QUALITATIVE RESEARCH METHODS (4 Credits)
An introduction to the theory and methods of qualitative research. Students will be exposed to various qualitative research methods through practical field exercises. These include ethnographic field observation, content analysis, interviewing, focus groups and unobtrusive measures. Other commonly used methods of collecting qualitative data are also examined.

Equivalent to: PPOL 523
Recommended: SOC 204 or SOC 204H

H 524. INTRODUCTION TO BIOSTATISTICS **(4 Credits)**
Quantitative analysis and interpretation of health data including probability distributions, estimation of effects, and hypothesis-tests such as Chi-square, one-way ANOVA, and simple linear regression.

H 564. COMPUTING TOOLS AND HEALTH DATA ANALYSIS **(3 credits)**
Modern computational biostatistics for analyzing health data, emphasizing important technologies and methods for data processing and understanding of how they work. Topics will evolve over time as new procedures are developed.

Corequisites: H 524

Genomics or high-performance computing

The Center for Genome Research and Biocomputing offers many short workshops, detailed at https://cgrb.oregonstate.edu/training/workshops

Many students will find these three introductory classes indispensable

Introduction to Unix/Linux (3 weeks @ 3 hrs per week)

This workshop introduces the natural environment of bioinformatics: the Linux command line. Material will cover logging into remote machines, filesystem organization and file manipulation, and installing and using software (including examples such as HMMER, BLAST, and MUSCLE).

Finally, we introduce the CGRB research infrastructure (including submitting batch jobs) and concepts for data analysis on the command line with tools such as grep and wc.

Command-Line Data Analysis (3 weeks @ 3 hrs per week)

The Linux command-line environment has long been used for analyzing text-based and scientific data, and there are a large number of tools pre-installed for data analysis. These can be chained together to form powerful pipelines. Material will cover these and related tools (including grep, sort, awk, sed, etc.) driven by examples of biological data in a problem-solving context that introduces programmatic thinking. This course also covers regular expressions, a useful syntax for matching and substituting string and sequence data.

Data Programming in R (6 weeks @ 3 hrs per week)

The R programming language is widely used for the analysis of statistical data sets. This course introduces the language from a computer science perspective, covering topics such as basic data types (e.g. integers, numerics, characters, vectors, lists, matrices, and data frames), importing and manipulating data (in particular, vector and data-frame indexing), control flow (loops, conditionals, and functions), and good practices for producing readable, reusable, and efficient R code. We'll also explore functional programming concepts and the powerful data manipulation and visualization packages dplyr and tidyr, and ggplot2.

Additional workshops on Genotyping-by-Sequencing, RNA-seq, Conservation Genomics and other topics will be of interest to many.

Introduction to Python I and II (each course is 3 weeks @ 3 hrs per week)

This course will introduce programming concepts driven by examples of biological data analysis. Topics covered include variables and data types (including strings, integers and floats, dictionaries and lists), control flow (loops, conditionals, and some boolean logic), functions, variable scope, variables as references (and nested structures), basic usage of regular expressions, and file I/O. The second half of this course expands on basic programming topics and explores best practices in software development and Object Oriented design, driven again by examples of biological data analysis.

In addition, several other departments offer relevant full classes.

IB 556. PHYLOGENETICS (4 Credits)

Explores the theory and practice of modern phylogenetic analysis. Emphasis placed on tree reconstruction algorithms, assessment of statistical support, and contemporary issues in phylogenetics. Lab will focus on the use of phylogenetic software and the analysis of molecular data sets. Lec/lab.

Equivalent to: BI 556

BOT 558. ECOSYSTEMS GENOMICS **(3 Credits)**
Genomic approaches used to understand species interactions with a focus on plant-associated microbes. Learning the conceptual framework and computational techniques of genomics to study the ecology of plant-microbe interactions at the ecosystem level.

Recommended: BI 311 and BI 314

BOT 560. FUNCTIONAL GENOMICS **(3 Credits)**
Functional genomics describes a set of conceptual approaches and associated laboratory techniques that rely on large-scale DNA sequence datasets to investigate the function of, and interactions between, genes as well as their RNA/protein products. This course will provide an overview of these techniques, including a) approaches to predicting protein function based on sequence analysis, b) large-scale genetic approaches to identifying novel genotype-phenotype associations, and c) transcriptomic, proteomic and metabolomic approaches that reveal gene functions by measuring changes in abundance/modification of associated RNA transcripts, proteins and metabolites.

Recommended: (BI 311 or BI 311H) and (BI 314 or BI 314H)

MCB 530. INTRODUCTION TO POPULATION GENETICS. (3 Credits)
Genetic polymorphisms, inbreeding, genetic drift, population subdivision and gene flow, mutation and selection. Emphasis on applied rather than theoretical questions. Offered alternate years.

Equivalent to: GEN 530
Recommended: BI 311 and ST 351 and ST 352

MCB 575. COMPARATIVE GENOMICS (4 Credits)
Principles of comparative genomics. Methods for genome assembly and annotation. Genomic approaches for the study of structural change, whole genome duplication, gene family evolution, gene networks, gene regulation and epigenetics. Lab topics include the analysis of next generation sequencing data and conducting comparative genome analyses. Lec/lab. CROSSLISTED as BOT 575.

Equivalent to: BOT 575
Recommended: Basic working knowledge of cell and molecular biology and genetics. BI 314 and (BI 311 or CSS 430)

MCB 576. INTRODUCTION TO COMPUTING IN THE LIFE SCIENCES **(3 Credits)**
Introduction to management of large datasets (e.g., nucleic acids, protein), computer programming languages, application of basic mathematical functions, and assembly of computational pipelines pertinent to life sciences. CROSSLISTED as BOT 476/BOT 576.

Equivalent to: BOT 576
Recommended: Cell and molecular biology or genetics and familiarity with text editing software and unix/linux operating system

Options for Students Who are Working Remotely and Need an Online ST Series See Ecampus Schedule of Classes for additional details:

https://ecampus.oregonstate.edu/soc/

ST 516 FOUNDATIONS OF DATA ANALYTICS (4)
Foundations of estimation and hypothesis testing; desirable properties of estimators; maximum likelihood; one- and two-sample problems; theoretical results are explored through simulations and analysis using R. Offered via Ecampus only. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 517 DATA ANALYTICS I (4)
Methods for modeling quantitative data and statistical learning--simple and multiple linear regression; linear mixed effects models; data imputation; prediction and cross-validation; scaling up to large datasets. Simulations and data analysis using R. Offered via Ecampus only. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 518 DATA ANALYTICS II (4)
Statistical methods and data analysis techniques for count data. Topics include tests for tables of counts, logistic regression, log-linear regression, generalized linear mixed models, and issues for large datasets. Data analysis in R. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 525 APPLIED SURVIVAL ANALYSIS (3)
Statistical methods for analyzing survival data or time-to-event data, which may be censored and/or truncated. Specific topics can vary term to term, and could include Kaplan-Meier estimator; K-sample hypothesis tests for survival data; Accelerated failure time model; Cox proportional hazard regression model. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 537 DATA VISUALIZATION (3)
Perceptual principles for displaying data; critique and improvement of data visualizations; use of color in visualization; principles of tidy data; strategies for data exploration; select special topics. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 539 SURVEY METHODS (3)
Survey design, data collection and analysis, general methodology. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 558 MULTIVARIATE ANALYTICS (3)
Basics of matrix algebra, principal components analysis, cluster analysis, factor analysis, multidimensional scaling. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.

ST 566 TIME SERIES ANALYTICS (3)
Focuses on statistical and analytical tools for analyzing data that are observed sequentially over time. Specific topics can vary term to term, and could include methods for exploratory time series analysis, linear time series models (ARMA, ARIMA), forecasting, spectral analysis and state-space models. The focus will be on applied problems, though some mathematical statistics is necessary for a solid understanding of the statistical issues. This course is designed for students in Data Analytics MS and Certificate programs. This course may be subject to Enforced Prerequisites that restrict registration into the course. Check the offerings below for more information.