**Statistical Computing Laboratory**

**Syllabus for Bios6605**

**Course Instructors:** Harry Smith, MPH

E-Mail: [Harry.Smith@ucdenver.edu](mailto:Harry.Smith@ucdenver.edu)

Caroline Ledbetter, MPH

E-Mail: [Caroline.Ledbetter@ucdenver.edu](mailto:Caroline.Ledbetter@ucdenver.edu)   
Office: Building 406, Office 210

**Canvas site: https://ucdenver.instructure.com/**

**Lab Meeting:** Wednesday, 10:30-11:20 AM

**Lab Locations with R:** P28-CTL 2201C, Ed 2 North or RC1 Room 1309

**Course Description:** The Statistical Computing Lab provides an introduction to R, a software package commonly used in health sciences and public health research. R assists in analyzing and interpreting data generated by implementing scientific research protocols. This computer-based statistical package is introduced to complement the learning of statistical concepts and to facilitate the interpretation and presentation of analytic results featured in basic Biostatistics courses.

The labs, examples, and assignments address interfacing with the statistical software package, managing research data, deriving descriptive statistics, using preprogrammed non-procedural statistical inference tools (procedures) and understanding and interpreting summary output. Some of the statistical procedures featured include: t-tests, chi-square tests, one-way ANOVA and simple linear and logistic regression. Data management skills are reviewed and practiced using the software’s interfaces to the computer’s native file system and other available internal commands and programming strategies that transform data for review and analyses, such as: sorting, ranking, recoding, combining, subsetting, filtering, reporting and archiving research data.

**Note:** This course has two parallels that use different software but have the same course objectives, instructional materials, and assignments. Students completing more than one of these programming courses (6603/6604/6605) may only be eligible to apply one credit hour toward their degrees. It is the student’s responsibility to confirm how these courses are applicable to his or her degree. Each course completed will be indicated on the student’s transcript but may not necessarily fulfill a graduation requirement.

Students can repeat this module any number of times and review any of the covered statistical packages at no additional cost and no additional degree credit. Modules within each software path are taught using the web-based, asynchronous course management and presentation software available through Canvas, website hosted at <https://ucdenver.instructure.com/>.

The targeted skill-set and critical tools are illustrated through review documents, examples in the fields of medicine, biology, epidemiology, and public health, and homework assignments that reinforce critical concepts in statistical analyses and interpretation of results. The primary goal of this course is to develop strategic statistical thinking by practicing rigorous and logical statistical programming. Students will be able to generate analytic results from sample datasets in this class, which can ultimately lead to rigorous written presentation and interpretation in other public health courses.

**Resource Needs:** To practice examples and complete assignments, students are frequently required to have access to the statistical software they wish to practice for several hours a week or more. R and R-Studio are free and easily downloaded and installed.

In addition, the school and campus have provided R access in the Ed-2 North computing lab (Room 2201C) and SAS in RC1 (Room 1309), which have been reserved at the times specified above. Approximately 30 systems have the statistical software packages installed. These can be used for any of the homework assignments. Additional times and/or location of labs will be posted as an Announcement on Canvas.

**Prerequisites:** BIOS6601 is a Co-requisite / Prerequisite for the Lab course. You must have a strong foundation in basic statistics in order to succeed in this course. The lab course assumes no previous knowledge or experience with either of the statistical packages or any other statistical software used for analyses and data management purposes. It is strongly recommended that the student have prior experience and knowledge of the Windows or Mac operating systems. (BIOS6680 is a SAS data management class that can prepare students to more thoroughly interact with the procedures and modules covered in these courses.)

**Course Objectives:** Student successfully completing this course should be able to:

* 1. Understand and be able to use the basic tools available in statistical packages,
  2. Be able to work effectively with analytical professionals who employ more advanced concepts and tools for data analysis;
  3. Be able to read, understand and judge the appropriateness of the use of biostatistics concepts and analytical tools in the public health, health sciences, and medical literature;
  4. Be comfortable with R software to analyze health-related data.
  5. Understand the logic and mechanics of statistical software for data management, data processing and statistical analysis, including:
* Accessing the main software interfaces to start the software, negotiate interactions within all of the sub-packages and interface with important operating sub-systems,
* Interacting with both internal and external files, including: text files containing both analysis commands and study data, Excel files housing study data, and other ODBC compliant data management software,
* Knowing how to move from the programming mode, to execution mode, to debugging mode, and to reading output generated by the system placed in an output window for storage, distribution and review.
* Save and print programs, logs, output file
* Save and archive analysis and raw data files.
* Know how to end the statistical session and maintain libraries of projects.
  1. Use the software for Data Management
  2. Use the software to input, modify and save data permanently and temporarily from either raw data sets or from the packages native file formats.
* Be able to use INPUT statements for text input,
* Be able to recode data using IF ... THEN ... ELSE type logic statements.
  1. Use several data processing steps within a program to sort, merge and create several files in one analysis file.
  2. Use the packages for Data Analysis
  3. Use the software to develop Probability distributions
  4. Use the software to perform random selection, random allocation and random variate generation using Uniform, Binomial, Poisson and Normal random number generators
  5. Use the software to generate simple descriptive statistics such as means, medians, ranges, standard deviations, standard errors and histograms for continuous data, and frequency distributions and percentages for categorical data
  6. Use various modules provided in the software to summarize and graph large data sets
  7. Use the software to perform hypothesis Tests – on Discrete and Continuous variables, for example: to perform one sample, two-sample and paired t-tests and be able to interpret the output
  8. Use the software for other purposes, e.g., to reorganize data and obtain ranks on a set of data,
  9. Use the software to perform Regression Analyses such as simple linear and multiple regression and interpret the output,
  10. Use the software to examine residuals, look for outliers, leverage and influence statistics, and obtain confidence intervals for parameters and predicted values,
  11. Use the software to compare two or more means and interpret its output,
  12. Use the software to perform Categorical data analysis, for example:
      1. logistic regression to test differences in Binomial proportions, interpret the coefficients, obtain odds ratios and confidence intervals, test hypotheses, and assess confounding and interaction.
      2. Mantel-Haenszel analysis for a case-control study and interpret the output,
  13. Use the software to perform Survival Analysis and to obtain Kaplan-Meier estimates, perform log-rank tests, and interpret other related output.

**Competencies to be developed:** At the completion of BIOS 6605, students will be able to:

1. Describe and demonstrate the role of statistical software to conduct research in public health and the health sciences.
2. Apply descriptive techniques commonly used to summarize public health data for public health evaluation and policy making.
3. Demonstrate key concepts of probability, random variation, and commonly used statistical probability distributions with simulations from statistical computing software.
4. Execute basic biostatistics concepts and appropriately apply statistical tools to the different measurement scales, and understand the implications of using inappropriate tools based on these distinctions.
5. Execute basic statistics in testing hypotheses and calculating confidence intervals and apply common statistical methods for inference.
6. Apply commonly used statistical tools when basic assumptions are not met.
7. Interpret results of statistical analysis found in the output of statistical procedures executed on public health studies.

**Integration of Statistical Computing Labs, Content and skills with other Core Areas:**

Every attempt will be made to use examples and scenarios that require students reflect on and relate statistical computing results to the areas and principles of: epidemiology, health behavior, environmental health, and health care organization and policy throughout the course.

# Reference Material:

**<https://cran.r-project.org/other-docs.html>**

[**http://swirlstats.com/**](http://swirlstats.com/)

**Canvas and Modes of Communications in the Class**

Course lecture handouts, Power Point slides examples and assignments are available through Canvas. To access the Statistical Lab site cut and paste or type

[**https://ucdenver.instructure.com/**](https://ucdenver.instructure.com/)into a web browser, login and then select the “Statistical Software Lab” to begin. Assignments can be submitted for grading using Canvas’s grading system.

Students with questions are encouraged to attend lab first. Students may email additional questions to the course TA or instructor as needed. Please keep in mind that programming questions are more difficult to answer via email. Many common questions will be addressed in lab.

**Required Lab Assignments**

Homework assignments will be provided each week during the course. Each week, students will have access to the current week’s assignment and the assignment for the following week. Assignments are mandatory activities for students since they constitute the mechanism by which students learn how to use statistical computing software and computing devices to perform statistical analysis. Since the assignments are computing exercises, there are often many reasonable options for completing the assignment. It is often (but not always) clear when a program works or not. So provided specific code is not considered the only correct answer but represents recommended approaches to analyzing research data. Suggested answers or approaches can be reviewed in review sessions.

**Lab Preparation**

Students should read the week’s Module prior to attending lab and completing each assignment.

**Lab Sessions**

Each week’s lab session will be dedicated to introducing concepts and answering questions relating to a designated assignment. The first 10-15 minutes of the lab session will be used to review statistical concepts, introduce programming concepts, and demonstrate the use of the program for the given assignment. During the remainder of the lab session, students may work on the week’s assignment and ask questions.

**Class Schedule: R**

*All assignments are due by 12 MIDNIGHT of the provided due date.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Day** | **Date** | **Agenda** | **INSTRUCTOR** |
| **Tuesday** | Aug 28th | **LAB:** Introduction to course | **Harry and Caroline** |
|  | Sep 4th | **DUE:** Nothing  **LAB:** Module 1: Reading and Saving Data | **Harry** |
|  | Sep 11th | **DUE:** HW 1  **LAB:** Module 2: Data Structures and Types | **Caroline** |
|  | Sep 18th | **DUE:** HW 2  **LAB:** Module 3: Random Number Generation | **Caroline** |
|  | Sep 25th | **DUE:** HW 3  **LAB:** Module 4: Random Number Allocation | **Harry** |
|  | Oct 2nd | **DUE:** HW 4  **LAB:** Module 5: Data Exploration | **Caroline** |
|  | Oct 9th | **DUE:** HW 5  **LAB:** Module 6: Factors | **Caroline** |
|  | Oct 16th | **DUE:** HW 6  **LAB:** Module 7: Contingency Tables | **Caroline** |
|  | Oct 23rd | **DUE:** HW 7  **LAB:** Module 8: Graphs and Plots Part 1 | **Caroline** |
|  | Oct 30th | **DUE:** HW 8  **LAB:** Module 9: Graphs and Plots Part 2 | **Caroline** |
|  | Nov 6th | **DUE:** HW 9  **LAB:** Module 10: T-tests | **Harry** |
|  | Nov 13th | **DUE:** HW 10  **LAB:** Module 11: Non parametric Tests | **Harry** |
|  | Nov 20th | **DUE:** HW 11  **LAB:** Module 12: Simple Linear | **Harry** |
|  | Nov 27th | **DUE:** Nothing  **LAB:** Module 13: Power and Sample Size | **Harry** |
|  | Dec 4th | **DUE:** HW 12  **LAB:** Module 14: Multiple Linear Regression | **Harry** |
|  | Dec 11th | **DUE:** Nothing  **LAB: REVIEW** | **Harry and Caroline** |
|  | Dec 18th | **NO CLASS**  **DUE:** BONUS HOMEWORK |  |

**Assignments**

Late assignments will not be accepted. All assignments should be submitted on Canvas in the form of a Word document. All required components must be included in the Word document. Raw code will not be accepted.

Assignment 9 is an optional assignment covering multiple regression. All students are encouraged to complete it, but it is not mandatory. Points earned on Assignment 9 will count as extra credit toward the final grade. The maximum number of points available on Assignment 9 is 12.5.

Be advised: Assignment 9 is considerably more involved and challenging than the other assignments. You are best off doing Assignments 1-8 and keeping up with the class.

**Student Evaluation**

Lab Assignments will be counted toward your final class grade. Assignment 9 is for extra credit. Assignments will be graded based on a pre-defined rubric. Scores will be viewable in Canvas, along with any comments provided by the instructors. If you have questions about your score, email the instructor for clarification. Since there are many programming solutions to a single problem, there may be a number of ways to earn full credit. Student progress in achieving the educational objectives of the course will be determined by completion and submission of the 8 computer programming assignments.

**Student Grading**

A final grade will be assigned by the following percentages.

|  |  |
| --- | --- |
| **Letter grade** | **%** |
| A | 100-93 |
| A- | 90-92.9 |
| B+ | 89.9-87 |
| B | 86.9-83 |
| B- | 82.9-80 |
| C+ | 79.9-77 |
| C | 76.9-73 |
| C- | 72.9-70 |
| D+ | 69.9-67 |
| D | 66.9-63 |
| D- | 62.9-60 |
| F | Below 60 |

**Course Policies:**

Make-up Policy

In the case of unexpected emergencies or unexpected events that make it impossible to complete assignment, let me know immediately by e-mail. Depending on the situation, we will work out a plan to complete the course. Typically, no assignments can be made up after the close of the semester, but some flexibility is allowed in certain situations.

Academic Conduct Policy

All students are expected to abide the Honor Code of the Colorado School of Public Health.  Unless otherwise instructed, all of your work in this course should represent completely independent work.  Students are expected to familiarize themselves with the Student Honor Code that can be found at

<http://www.ucdenver.edu/academics/colleges/PublicHealth/Academics/academics/Documents/PoliciesHandbooks/CSPH_Honor_Code.pdf>  or the Student Resources Section of the CSPH website.  Any student found to have committed acts of misconduct (including, but not limited to cheating, plagiarism, misconduct of research, breach of confidentiality, or illegal or unlawful acts) will be subject to the procedures outlined in the CSPH Honor Code.

Special Needs

If you have special needs and wish to request accommodations, please contact the Office of Disability Resources and Services (DRS) located in Building 500, Room W1103.  DRS staff will assist in determining reasonable accommodations as well as coordinating the approved accommodations.  Contact information:  Phone:  303-724-5640; Fax:  303-724-5641; Email:  [sherry.holden@cudenver.edu](mailto:sherry.holden@cudenver.edu).  Please also talk with the instructor about any needs or concerns.

Course Times and Proctored Sessions

Class attendance is not mandatory but interactions with Canvas, the instructor and fellow students are highly encouraged for sharing knowledge and experiences. Attendance at each individual group session is optional since these are self-paced modules that can be done on systems with a web browser and the appropriate statistical software. Students can attend any of the lab sessions on a drop-in basis when extra help or clarification if needed, however, if all the computers are in use you may have to share a system or use your own laptop.

Purchasing Statistical Software

R is free and can be downloaded from https://www.r-project.org/.