

**UNIVERSITY OF MASSACHUSETTS**  
**Department of**  
**Biostatistics and Epidemiology**  
**BIOSTATS 690C – Data Management & Applied Data Analysis with Stata (and R too!)**  
**Fall 2020**

***New! Fall 2020***

Dear BIOSTATS 690C Fall 2020,  
 This year, BIOSTATS 690C, in **addition to Stata**, will also an introduction to **RStudio**.  
 Please be aware that this is a “work in progress” as I transition this course from Stata- to R-based.  
 Thank you – Carol

<b>Title:</b>	BIOSTATS 690C – Data Management & Applied Data Analysis
<b>Course website:</b>	<a href="https://people.umass.edu/~biostat690c/">https://people.umass.edu/~biostat690c/</a>
<b>Class number:</b>	67964
<b>Number of credits:</b>	3
<b>Instructor:</b>	Carol Bigelow, PhD Department of Biostatistics & Epidemiology Email: <a href="mailto:cbigelow@schoolph.umass.edu">cbigelow@schoolph.umass.edu</a>
<b>Teaching Assistant:</b>	Aaron Gerding, Graduate Student - Biostatistics Email: <i>available upon request to</i> <a href="mailto:cbigelow@schoolph.umass.edu">cbigelow@schoolph.umass.edu</a>
<b>Zoom classes (synchronous):</b>	Available upon request, with dates and times by mutual agreement
<b>Office hours (synchronous):</b>	Thursdays 4:00 – 5:00 pm EDT ( <i>EST beginning November 1, 2020</i> )

Course Description .....	2
ADA Accommodation Policy .....	2
Covid-19 Impact on this Course .....	2
Policies on Classes and Work Missed for Extenuating Circumstances .....	3
Getting Started with Remote Learning .....	3
Course Units, Objectives and Outcome Competencies .....	4
Textbook, Software, and Internet Resources .....	5
Important Dates to Remember .....	5
Course Schedule (Units, Homeworks, Exams) .....	6
Course Expectations (Instructor, Teaching Assistants, and Students) .....	7
Grade Determination .....	8
Policy on Late Submissions .....	8
Policy on Academic Dishonesty .....	9
Valuing, Recognizing, and Encouraging Diversity .....	9
Names and Pronouns .....	9
Title IX Statement .....	10
Copyright Protection .....	10

## Course Description:

This course is an introduction to the design, management, and use of data management systems for the collection and analysis of research data, especially epidemiologic research data on humans. MS Excel and Stata are emphasized. Topics include data base development, Health Insurance Portability Accountability Act (HIPAA) compliance, data manipulation and cleaning, data summarization, and selected topics in statistical analysis programming. This course is designed for the UMass SPHHS student who has completed [BIOSTATS 540 – Introduction to Biostatistics](#) and [BIOSTATS 640 – Intermediate Biostatistics](#) and who seeks training in data management and the use of STATA version 16. **Plus, beginning Fall 2020, an introduction to R and RStudio is also being provided.**

Prerequisites: BIOSTATS 540 and BIOSTATS 640; or permission of instructor.

## ADA Accommodation Policy

The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier-free campus. If you have a disability and require accommodations, please register with Disability Services (161 Whitmore Administration building; phone 413-545-0892) to have an accommodation letter sent to your faculty. Information on services and materials for registering are also available on their website [www.umass.edu/disability](http://www.umass.edu/disability).”

If, because of a disability, you may require special arrangements in order to meet course requirements, please contact me as soon as possible (email: [cbigelow@schoolph.umass.edu](mailto:cbigelow@schoolph.umass.edu)) and we will make the necessary arrangements.

## COVID-19 Impact on this Course

The ongoing COVID-19 pandemic presents new challenges that can be stressful for any of us. In Fall 2020, I will teach this course using remote learning and this means that we will be engaging both synchronously and asynchronously. Here are some key things to know about how this course will run:

Asynchronous Learning. Because BIOSTATS 690C is an “online” class, most of your engagement with this course will be asynchronous. Each week, I will provide course materials, course activities and online resources for that week’s topic. These will be posted in two (2) locations: the course website (<https://people.umass.edu/~biostat690c/>) and in Blackboard Learn.

***Tip – It sometimes happens that Blackboard Learn is down or unavailable for some reason. When this occurs, you can always find what you need on the course website.***

- **Synchronous Learning.** Again, because this is an “online” class, there is no scheduled meeting time. However, depending on interest, and upon request for particular face-to-face meetings, I would be happy to schedule selected Zoom meetings. Keep this in mind and let me know what you think!
- All Zoom meetings will be recorded.
- **Important - I will correspond with you using your UMail email address only.**

Please be aware that we may need to make adjustments as the pandemic develops throughout the semester. I will do my best to be flexible and accommodate unforeseen needs. Please reach out ([cbigelow@schoolph.umass.edu](mailto:cbigelow@schoolph.umass.edu)) if you have any concerns, needs, or suggestions.

### **Policies on Classes and Work Missed for Extenuating Circumstances.**

Per University of Massachusetts Academic Regulations, “*Students absent due to extenuating circumstances-including jury duty, military obligations, scheduled activities for other classes, the death of a family member, or verifiable health-related incapacity-remain responsible for meeting all class requirements and contacting the faculty member in a timely fashion about making up missed work. Faculty shall offer such students reasonable assistance in making up missed classes (i.e., making arrangements for attendance at labs or discussion sections which meet at other times; providing makeup exams or labs where feasible or offer mutually agreeable alternatives to make up work).*”

If any extenuating circumstances prevent you from completing any work, please contact me as soon as possible (email: [cbigelow@schoolph.umass.edu](mailto:cbigelow@schoolph.umass.edu)) so that we can make alternative arrangements.

### **Getting Started with Remote Learning**

Please be sure to familiarize yourself with these resources and technologies before the first week of class (August 24-28, 2020).

1. Public Course Website: <https://people.umass.edu/~biostat690c/>
2. UMass Amherst Blackboard Learn: <https://uma.umassonline.net/webapps/login/#>  
Resource for Learning Blackboard Learn:  
(source: Rowan University, Campbell Library) [Blackboard Tutorials for Students](#)
3. UMass Amherst Zoom Login: <https://www.umass.edu/it/zoom>  
Resource for Learning Zoom:  
(source: <https://support.zoom.us>) [Zoom Video Tutorials](#)  
**Tip:** Be sure to watch these two tutorials, here: “[Join a Meeting](#)” and “[Meeting Controls](#)”

## Course Units, Objectives, and Outcome Competencies

**Course Units.** This course has 10 units.

1. Principles of Data Management
2. Ethical Management of Human Subjects Information
3. MS Excel for Epidemiology
4. Introduction to STATA (or RStudio)
5. STATA (or RStudio) for Data Description
6. STATA (or RStudio) for Graphs
7. STATA (or RStudio) for Analysis of 1 and 2 (or more) Samples
8. STATA (or RStudio) for Categorical Data
9. STATA (or RStudio) for Normal Theory Regression
10. STATA (or RStudio) for Logistic Regression

**Objectives.** The goal of this course is to develop a working knowledge of data manipulation, best practices for the ethical maintenance of human subjects' data, data summarization and its communication to lay and professional audiences, and the use of Stata (or RStudio) for basic data manipulations and statistical analysis.

**Outcome Competencies.** Specific outcome competencies that are among the goals of this course include, but are not limited:

1. Best practices for the ethical maintenance of human subjects' data – Students will gain knowledge of the Health Insurance Portability Accountability Act (HIPAA) and will be expected to incorporate its understanding into the structure of their data management system.
2. The preparation of data that is ready for statistical analysis – Students will gain experience in several steps related to compiling data that is ready for statistical analysis: error checking, audit trail creation, formatting, labeling, etc. Students will also become knowledgeable in merging files.
3. The organization and documentation of a data management system – Each student will be expected to design and document a data management system that includes, but is not limited to, specification of directory structure, file naming, and back-up.
4. Data Summarization for presentation to professional and lay audiences – Students will gain experience in using MS Word, MS Excel, and Stata (or RStudio) for the production of tabular and graphical summaries of public health data. Students will also be expected to write brief descriptions of these summaries. **Note – in this course, we will not be using MS Excel in data analyses.**
5. Stata (or RStudio) for basic statistical analysis of epidemiologic data – Students will acquire experience in using Stata for basic statistical analyses of public health data: one and two samples, analysis of variance, multiple linear regression, stratified analysis of rates, and logistic regression. Time permitting, students will gain experience using Stata (or RStudio) for other analyses of their choosing.

## Textbook, Software, and Internet Resources

### **Textbook.**

Acock, Alan C.

*A Gentle Introduction to Stata, Sixth Edition.* (Note – older editions are just fine!)

Stata Press, 2018.

ISBN-13: 978-1-59718-269-0

### **Software.**

We will be using the following software:

1. MS Excel
2. Stata version 16 (If you already have an older version, it is **NOT** necessary to upgrade)
3. (optional) R and RStudio (we will actually be doing all our work in RStudio)
4. And also: MS Word

**You do NOT need to have obtained any statistical software before the class starts. I will provide you with complete instructions in the first weeks.**

### **Internet Resources.**

I will be providing lots of internet resources for learning Stata and R/RStudio as we go along. Here are a couple of good sites you might want to bookmark!

For Stata Learners.

(source: Stata Corp, LLC) <https://www.stata.com/links/resources-for-learning-stata/>

For R Learners

(source: Melinda Higgins, Emory University) <https://melindahiggins2000.github.io/N741bigdata/help.html>

## Important Dates to Remember

First Week of Class	Monday – Friday August 24-28, 2020
Last Day to Drop with no Record	Monday September 7, 2020
Last Day to Drop with “DR”	Friday October 16, 2020
Last Week of Class	Monday-Friday November 16-20, 2020
Test 3 (Units 8-10) Due	Wednesday December 6, 2020

## Course Schedule (Units, Homeworks, and Exams)

### Please note

I will try to post materials on Mondays but cannot guarantee that I will always be able to do this.

### Units and Homeworks.

Week	Date	Unit	Homework Posting and Due Dates
1	Monday - Friday August 24-28, 2020	Up and Running with Stata and R, RStudio!	No homework this week
2	Wednesday - Tuesday September 2-8, 2020	1 – Principles of Data Management 2 – Ethical Management of Info	Posted: Wednesday September 2, 2020 Due: Wednesday September 9, 2020
3	Wednesday - Tuesday September 9-15, 2020	3 – MS Excel	Posted: Wednesday September 9, 2020 Due: Wednesday September 16, 2020
4	Wednesday - Tuesday September 16-22, 2020	4 – Introduction to Stata/RStudio Part 1 of 2	Posted: Wednesday September 16 2020 Due: Wednesday September 30, 2020
5	Wednesday - Tuesday September 23-29, 2020	4 – Introduction to Stata/RStudio Part 2 of 2	No new homework this week
6	Wednesday - Tuesday September 30 – October 6, 2020	5 – Stata/RStudio for Data Description	Posted: Wednesday September 30 2020 Due: Wednesday October 7, 2020
7	Wednesday - Tuesday October 7-13, 2020	6 – Stata/RStudio for Graphs	Posted: Wednesday October 7, 2020 Due: Wednesday October 14, 2020
8	Wednesday - Tuesday October 14-20, 2020	7 – Stata/RStudio for 1, 2, 3+ Samples	Posted: Wednesday October 14, 2020 Due: Wednesday October 21, 2020
9	Wednesday - Tuesday October 21-27, 2020	8 – Stata/RStudio for Categorical Data	Posted: Wednesday October 21, 2020 Due: Wednesday October 28, 2020
10	Wednesday - Tuesday October 28 – November 3, 2020	9 – Stata/RStudio for Normal Theory Regression. Part 1 of 2	Posted: Wednesday October 28, 2020 Due: Wednesday November 11, 2020
11	Wednesday - Tuesday November 4-10, 2020	9 – Stata/RStudio for Normal Theory Regression. Part 2 of 2	No new homework this week
12	Wednesday - Tuesday November 11-17, 2020	10 – Stata/RStudio for Logistic Regression. Part 1 of 2.	Posted: Wednesday November 11 2020 Due: Wednesday December 6 2020
13	Wednesday - Friday November 18-20, 2020	10 – Stata/RStudio for Logistic Regression. Part 2 of 2.	No new homework this week
-	COURSE CLOSEOUT		Wednesday December 6, 2020 EXAM 3 DUE Today

### Tests.

	<b>Posted</b>	<b>Due Date</b> <b>(Last date for credit)</b>
<b>Test 1</b> Unit 3 (MS Excel) Unit 4 (Introduction to Stata/RStudio)	<i>Wednesday</i> September 23, 2020	<i>Wednesday</i> <b>October 7, 2020</b>  <u>Last Date for Submission</u> (-10 points) Friday October 9, 2020  (-20 points) Wednesday October 14, 2020
<b>Test 2</b> Unit 4 (Introduction to Stata/RStudio) Unit 5 (Stata/RStudio for Data Description) Unit 6 (Stata/RStudio for Graphs) Unit 7 (Stata/RStudio for 1, 2, 3+ Samples)	<i>Wednesday</i> October 28, 2020	<i>Wednesday</i> <b>November 11, 2020</b>  <u>Last Date for Submission</u> (-10 points) Friday November 13, 2020  (-20 points) Wednesday November 18, 2020
<b>Test 3</b> Unit 8 (Stata/RStudio for Categorical Data) Unit 9 (Stata/RStudio for Regression) Unit 10 (Stata/RStudio for Logistic Regression)	<i>Wednesday</i> November 18, 2020	<i>Wednesday</i> <b>December 6, 2020</b>  <u>Last Date for Submission</u> (-10 points) Friday December 8, 2020

## Course Expectations (Instructor, Teaching Assistants, and Students)

### Instructor

- I will respond to all emails and Blackboard messages every day, *except for Saturdays*
- I will provide regularly scheduled Zoom office hours on Thursdays 4:00 – 5:00 pm
- I will also provide Zoom meeting office hours by appointment, as appropriate.

### Teaching Assistant (TA)

- TAs will respond to Blackboard discussion thread questions every day, *except for Saturdays*
- TAs will record homework submissions within one week of their due dates.
- TAs will respond to questions in homework submissions within one week of their due dates

## Students

- Attendance at Zoom classes is **not** required
- Students will abide by the University of Massachusetts policy on academic dishonesty
- Students will abide by the policy on late submissions (see below)
- To earn full credit on the homework, students must submit 8 homework assignments

## Grade Determination

### Course Score determination

	Percent of Grade
<b>Homeworks</b> Homeworks are graded pass/fail and the solutions are provided (and you are welcome to consult the solutions as you go along!). Thus, the homeworks are the participation portion of your grade.  <u>Late policy.</u> If you submit your work on the due date, your score will be 100. Late homework submissions are accepted, per the late policy described below. Late homeworks submitted within 48 hours will be given a score of 90. Homeworks submitted after 48 hours late, but no later than one week, will be given a score of 80. Homeworks submitted more than one week late will be given a score of 0.  To earn full homework credit, you must complete eight (8) assignments.	<b>25%, <u>sub-total</u></b>
<b>Exams (all open book)</b> There are 3 tests, all open book. For each test, you are welcome to consult any resource you like, but <b><i>you are <u>not</u> allowed to consult any person except me.</i></b>  <u>Late policy</u> If you submit your exam on the due date, you will earn full credit for your work. Exams submitted late but within 48 hours will have 10 points deducted from their score. Exams submitted late but between 48 hours and 1 week will have 20 points deducted from their score. Exams submitted more than one week late will be given a score of 0.	<b>75%, <u>sub-total</u></b>  Best test – 40% 2 <sup>nd</sup> best test – 20% 3 <sup>rd</sup> best test – 15%

**Letter grade determination.** Your course score will be converted to a letter grade as follows:

A	95 and over
A-	90 - 94
B+	87 - 89
B	83 – 86
B-	80 - 82
C+	77 – 79
C	70 – 76
F	Below 70



### Policy on Late Submissions

These are challenging times and so, very possibly, as this course progresses a deadline will pose an unreasonable burden. *Therefore, this course has a policy on late submissions that aims to be fair while at the same time accommodating those who, for whatever reason, need an extension.*

	Credit Policy - Homeworks	Grading Policy - Exams
On Time	Full Credit for points Scored	Full Credit for points Scored
Up to 48 hours late	Full Credit for points Scored	Points Scored – 10 points
3-7 days late	Points Scored – 20 points	Points Scored – 20 points
8+ days late	0 points (no credit)	0 points (no credit)

### Policy on Academic Dishonesty

All students are expected to adhere to guidelines of University of Massachusetts regarding academic honesty. A copy of these guidelines is available online at

[www.umass.edu/dean\\_students/code\\_conduct/acad\\_honest.htm](http://www.umass.edu/dean_students/code_conduct/acad_honest.htm)

The University of Massachusetts/Amherst Senate Document 89-026 defines academic dishonesty as including but not limited to:

- Cheating – intentional deceit, trickery, or breach of confidence, used to gain some unfair or dishonest advantage in one’s academic work.
- Fabrication – intentional falsification or invention of any information or citation in any academic exercise.
- Facilitating dishonesty – knowingly helping or attempting to help someone else commit an act of academic dishonesty.
- Plagiarism – knowingly representing the words or ideas of another as one’s own work in any academic exercise.
- Submitting in whole or in part, without citation, prewritten term papers of another or the research of another (including but not limited to such materials sold or distributed commercially).

### Valuing, Recognizing, and Encouraging Diversity

I believe that promoting and valuing diversity in the classroom enriches learning and broadens everyone’s perspectives. I also believe in inclusion, tolerance and respect for others as essential values. Where possible, I will strive to create a sense of community and promote excellence in the learning environment. With respect to diversity, I will seek out and honor (1) the variety of life experiences you have had, and (2) the factors that define your “diversity of presence,” including: age, economic circumstances, ethnic identification, disability, gender, geographic origin, race, religion, sexual orientation, social position.

### Names and Pronouns

If you have not already indicated your chosen first name and pronouns in SPIRE, please let me know what name and pronouns we should use for you (email: [cbigelow@schoolph.umass.edu](mailto:cbigelow@schoolph.umass.edu)).

## Title IX Statement

The University of Massachusetts Amherst is committed to fostering a safe, productive learning environment. Title IX and our school policy prohibits discrimination on the basis of sex. Sexual misconduct — including harassment, domestic and dating violence, sexual assault, and stalking — is also prohibited at our school.

UMass Amherst encourages anyone experiencing sexual misconduct to talk to someone about what happened, so they can get the support they need and our school can respond appropriately. If you wish to speak confidentially about an incident of sexual misconduct, want more information about filing a report, or have questions about school policies and procedures, please contact our Title IX Coordinator, Débora D. Ferreira, Equal Opportunity Office (EO), 413-545-3464, [equalopportunity@admin.umass.edu](mailto:equalopportunity@admin.umass.edu).

Please be aware. UMass Amherst is legally obligated to investigate reports of sexual misconduct, and therefore it cannot guarantee the confidentiality of a report, but it will consider a request for confidentiality and respect it to the extent possible. If you want to talk with someone who is not a mandated reporter, you can contact the Center for Women and Community, (<https://www.umass.edu/cwc/>, 413-545-0883, or 24-hour hotline 413-545-0800), the Center for Counseling and Psychological Help (<https://www.umass.edu/counseling/>, 413-545-2337), or University Health Services SANE program (<https://www.umass.edu/uhs/services/sane>, 413-577-5000). Please also be aware. As an instructor, I am also required by our school to report incidents of sexual misconduct and thus cannot guarantee confidentiality. I must provide our Title IX coordinator with relevant details such as the names of those involved in the incident.

## Copyright Protection

Many of the materials created for this course are my own intellectual property. This includes, but is not limited to the syllabus, lectures, and course notes. Except to the extent not protected by copyright law, any use, distribution or sale of such materials requires my permission. Please be aware that it is a violation of university policy to reproduce, for distribution or sale, class lectures or class notes, unless the faculty member has explicitly waived copyright.