Mel and Enid Zuckerman College of Public Health
University of Arizona

SYLLABUS
Biostatistics 576B Biostatistics for Research
SPRING 2021
3 Units

Time: Monday and Wednesday 10:30 am – 11:50 am

Location: Lectures: Zoom online
Drachman A114 (in person when feasible)

Labs: Zoom online (STATA) – Shen Liu
Zoom online (SAS/R) – Xingyi Yang

Instructor: Denise J. Roe, Dr.P.H.
Professor, Epidemiology & Biostatistics
1933 University of Arizona Cancer Center
Telephone: (520) 626-2281
droe@email.arizona.edu

Office Hours: Zoom online
Monday and Wednesday 3:00 pm – 4:00 pm

By appointment (email to schedule an appointment):
Directions to 1933 University of Arizona Cancer Center:
Enter the Cancer Center using the doors closest to BUMC
(south-west corner of the building)
At the end of the hallway turn left
Take the stairwell on the right down to the first floor
Exit the stairwell and turn left
Take the first right
Follow the signs to the Biometry Shared Service
My office is the next to last door on the right

Laboratory Instructors: Shen Liu shenliu@email.arizona.edu
Xingyi Yang xingyiyan92@email.arizona.edu
Lab Instructor Office Hours:

<table>
<thead>
<tr>
<th>Day</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>7:00 pm – 9:00 pm (Shen Liu)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>7:00 pm – 9:00 pm (Xingyi Yang)</td>
</tr>
</tbody>
</table>

Catalog Description: Descriptive statistics and statistical inference relevant to biomedical research, including data analysis, regression and correlation analysis, analysis of variance, survival analysis, statistical methods for epidemiology and statistical evaluation of clinical literature.

Course Description: An intermediate course in biostatistical methods and their application in epidemiology and public health. This course covers linear regression analysis, logistic regression analysis, time-to-event (survival) analysis, and study design.

Course Prerequisites

Biostatistics: Biostatistics 576A (Introduction to Biostatistics)

Note: If you did not take Biostatistics 576A at UA please see me to make sure that you have the necessary prerequisites for the course

Epidemiology: Epidemiology 573A (Basic Principles in Epidemiology)

Computing: Ability to use a statistical computing package (STATA, SAS, R)

Course Objectives: During the course, students will:

- Apply statistical principles of sample size and power estimation to aid in the design of studies

- Use statistical modeling techniques, such as linear regression, logistic regression, log-binomial regression, time-to-event analysis, and Poisson regression, as appropriate, to meet the scientific objectives in research studies

- Analyze data by deciding the appropriate statistical techniques, ensuring that the assumptions are met, and effectively communicating analytic results

- Critically evaluate the use of statistics in published journal articles
Learning Outcomes (Competencies Obtained):

Program Competencies Covered (MPH Program Level):

Upon completion of the course, students will be able to:

Evidence-based Approaches to Public Health:
1. Apply epidemiological methods to the breadth of settings and situations in public health practice
2. Select quantitative and qualitative data collection methods appropriate for a given public health context
3. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate
4. Interpret results of data analysis for public health research, policy or practice

Communication:
19. Communicate audience-appropriate public health content, both in writing and through oral presentation

Biostatistics Concentration Competencies Covered (MPH/MS/PhD Programs):

Upon completion of the course, students will be able to:

Biostatistics MPH Competencies Covered:
1. Utilize appropriate statistical tools to address specific scientific questions.
2. Explain statistical concepts and findings to a general scientific audience.

Biostatistics MS Competencies Covered:
1. To recognize strengths and weaknesses of proposed statistical approaches, including alternative designs, data sources, and analytical methods.
2. To suggest preferred methodological alternatives to commonly used statistical methods when assumptions are not met.

Biostatistics PhD Competencies Covered:
1. Apply descriptive and inferential methodologies according to the type of study design for answering a particular research question
2. Communicate understanding of the assumptions necessary for a given statistical procedure as well as the ability to determine if the assumptions are met for a given study design or data set

Course Notes: A website has been created for this class using the Desire 2 Learn (D2L) interface. The course website contains the syllabus, class notes, Zoom recordings, Panopto recordings (if applicable) and homework assignments.

Class announcements also will be posted on this site, so it is a good idea to check the site before each class to stay current.

STATA and SAS will be used throughout the course. For each lecture two versions are provided – one using STATA and the other using SAS. The material will be the same in each format. Examples using R also are available.
To access the 576B website, login at: [http://d2l.arizona.edu](http://d2l.arizona.edu)

- Click the ‘UA NetID’ Login.
- Enter your NetID and password, as you would to access your UA email account.
- Under ‘My Courses’, click on: ‘BIOS 576B SP21 001 101 - 102 202’
  - Announcements: This section contains any class announcements, such as the material to be covered in each class and suggestions/hints for the homework.
  - Content: Access the syllabus, Zoom and Panopto recordings, class notes, homework assignments and supplemental information in this section.

For further information on how to use the D2L interface, go to: [http://www.help.d2l.arizona.edu](http://www.help.d2l.arizona.edu)

Note that if you do not have a UA NetID, please see me so that I can give you access to the D2L site.

To access the Panopto recorded lecture you must use Firefox or Chrome as the browser when you log into D2L. Internet Explorer will not work.

To access a recorded lecture click on the “Lecture Recordings” module. Then “Click here” to go to our class recordings.

**Recommended Texts**


The 8th Edition is available for purchase in the AHSC bookstore or from the publisher or other online sites. Alternatively you may rent a hard copy or purchase electronic access from the publisher. The link is:


The link for the book companion site is:


The book companion site includes:
- Data Set Descriptions (.doc)
- Data Sets (SAS and STATA)
- Study Guide (includes a summary of each chapter with additional problems and solutions)


(* below implies that the material is more advanced than required for the course)

Course Requirements:

Homework

- 8 graded homework assignments
- Questions will be answered about the homework during the lab sessions (Mondays)
- To be graded they must be turned in by the due date
- **Late homework assignments will not be accepted**

- Keep copies of all of your homework so that you can study for the exams while your submitted work is being graded by the TA.
- You can drop your **lowest** homework score. It is best to reserve this for times that you are unexpectedly out of town, cannot turn in the homework due to illness, or your computer crashes with your homework on it.

Examinations

<table>
<thead>
<tr>
<th>Exam</th>
<th>Date Distributed</th>
<th>Date Due</th>
<th>Lectures and homework</th>
</tr>
</thead>
</table>
| Midterm 1| Wednesday, February 17 | Tuesday, February 23 | LI - 1 through LI – 6  
CA – 1 and CA - 2  
Homework 1 – 3 |
| Midterm 2| Wednesday, March 31  | Tuesday, April 6  | Cumulative with emphasis on  
CA - 3 through CA – 4  
LO – 1 through LO – 6  
Homework 4 - 6 |
| Final    | Wednesday, May 5   | Tuesday, May 11  | Cumulative with emphasis on  
SU-1 through SU – 7  
Homework 7 - 8 |

- All exams will be take-home. They will include a dataset with scientific questions to be addressed and interpretation of results from published papers.
- Exams must be turned in no later than midnight on the date due to the appropriate D2L assignment tab.
- For each exam, all students must sign the following statement (first page of the exam):
I have not discussed any aspects of this exam with other class members, former class members, other students, or faculty. I understand that if there is evidence that I have violated these restrictions, my grade on the exam will be reduced by 50%.

Signature  
Printed Name  
Date

Extra Credit

- **No** extra credit is available for the course

Grading Scale/Student Evaluation and Policies:

<table>
<thead>
<tr>
<th>Assessment Methods</th>
<th>Due Dates</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>See course schedule below</td>
<td>100 (each assignment weighted the same, even if length differs)</td>
</tr>
<tr>
<td>Midterm 1</td>
<td>February 23</td>
<td>300</td>
</tr>
<tr>
<td>Midterm 2</td>
<td>April 6</td>
<td>300</td>
</tr>
<tr>
<td>Final</td>
<td>May 11</td>
<td>300</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1000</strong></td>
</tr>
</tbody>
</table>

Description of each Assessment and Competencies Covered by the Assessment

**Homework:** Homework assignments build statistical analysis and interpretation skills

**Midterms 1 and 2 and Final:** MPH Competencies and Biostatistics Competencies listed above

Final grades are based on the following point system:

- A = 90-100%
- B = 80-89%
- C = 70-79%
- D = 60-69%
- E = 59% or less

The instructor reserves the right to revise this scale, if necessary.
**Course Schedule**: Any changes to the following schedule will be announced in lecture or the D2L site. You are responsible for obtaining information on any changes.

<table>
<thead>
<tr>
<th>Notes</th>
<th>Date</th>
<th>Topic</th>
<th>Homework Due</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Linear Regression Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overview</td>
<td>LI – 1</td>
<td>Overview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed, 1/13</td>
<td></td>
<td>Review of multiple linear regression</td>
<td>R 11.9</td>
<td>V 4.1 – 4.3.3</td>
</tr>
<tr>
<td>Mon, 1/18</td>
<td></td>
<td>No class – Martin Luther King Day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI – 2</td>
<td>Wed, 1/20</td>
<td>Regression diagnostics</td>
<td>V 4.7</td>
<td></td>
</tr>
<tr>
<td>Mon, 1/25</td>
<td></td>
<td><em>Homework 1 lab</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI – 3</td>
<td>Wed, 1/27</td>
<td>Analysis of covariance</td>
<td>Homework 1</td>
<td>A 9.3, V 4.6</td>
</tr>
<tr>
<td>LI – 4</td>
<td>Wed, 1/27</td>
<td>Predictor selection methods</td>
<td></td>
<td>A Ch 8</td>
</tr>
<tr>
<td>Mon, 2/1</td>
<td></td>
<td><em>Homework 2 lab</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI – 5</td>
<td>Wed, 2/3</td>
<td>Modeling strategies</td>
<td>Homework 2</td>
<td>V 10.1 – 10.6</td>
</tr>
<tr>
<td>LI – 6</td>
<td>Wed, 2/3</td>
<td>Missing data and imputation</td>
<td></td>
<td>A 9.2, V 11.1 – 11.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Categorical Data Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA – 1</td>
<td>Mon, 2/8</td>
<td>Hypothesis testing and power for a 2 x 2</td>
<td>R 10.5</td>
<td></td>
</tr>
<tr>
<td>CA - 2</td>
<td></td>
<td>table</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Power and sample size estimation in</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>epidemiology studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA – 3</td>
<td>Wed, 2/10</td>
<td>Stratified analysis</td>
<td>R 13.6</td>
<td></td>
</tr>
<tr>
<td>Mon, 2/15</td>
<td></td>
<td><em>Homework 3 lab</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed, 2/17</td>
<td></td>
<td>Review</td>
<td>Homework 3</td>
<td></td>
</tr>
<tr>
<td>Mon, 2/22</td>
<td></td>
<td><em>Midterm 1 Distributed</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tues, 2/23</td>
<td></td>
<td>No class</td>
<td>By Midnight</td>
<td></td>
</tr>
<tr>
<td>CA – 4</td>
<td>Wed, 2/24</td>
<td>Stratified Analysis: Tests for trend</td>
<td>R 13.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO – 1</td>
<td>Mon, 3/1</td>
<td>Logistic regression</td>
<td>R 13.8</td>
<td>V 5.1 – 5.2.1</td>
</tr>
<tr>
<td>LO – 2</td>
<td>Wed, 3/3</td>
<td>Review Midterm 1</td>
<td>V 5.2.2 – 5.2.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic regression: Interaction and</td>
<td></td>
<td>H 3.6 – 3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>confounding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td>Date</td>
<td>Topic</td>
<td>Homework Due</td>
<td>Readings</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>--------------------------------------------</td>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Mon, 3/8</td>
<td></td>
<td>Homework 4 lab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed, 3/10</td>
<td></td>
<td>Reading Day No class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mon, 3/15</td>
<td></td>
<td>Homework 5 lab</td>
<td>Homework 4</td>
<td></td>
</tr>
<tr>
<td>LO – 3</td>
<td>Wed, 3/17</td>
<td>Logistic regression: Goodness-of-fit</td>
<td>Homework 5</td>
<td>V 5.2.5 – 5.2.6, 5.4.3.2</td>
</tr>
<tr>
<td>LO – 4</td>
<td></td>
<td>Logistic regression: Diagnostics</td>
<td></td>
<td>H 5.1 – 5.2</td>
</tr>
<tr>
<td>Mon, 3/22</td>
<td></td>
<td>Homework 6 lab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO – 5</td>
<td>Wed, 3/24</td>
<td>Multinomial and ordinal logistic regression</td>
<td>Homework 6</td>
<td>V 5.5.6</td>
</tr>
<tr>
<td>LO – 6</td>
<td>Mon, 3/29</td>
<td>Log-binomial regression Poisson regression</td>
<td>V 5.5.3, H 5.3</td>
<td></td>
</tr>
<tr>
<td>LO – 7</td>
<td></td>
<td></td>
<td></td>
<td>V 8.1</td>
</tr>
<tr>
<td>Wed, 3/31</td>
<td></td>
<td>Review Midterm 2 Distributed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mon, 4/5</td>
<td></td>
<td>No class</td>
<td>By Midnight</td>
<td></td>
</tr>
<tr>
<td>Tues, 4/6</td>
<td></td>
<td>Midterm 2 Due</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Survival Analysis**

<table>
<thead>
<tr>
<th>SU – 1</th>
<th>Wed, 4/7</th>
<th>Survival Analysis</th>
<th>V 3.5.1 – 3.5.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SU – 2</td>
<td></td>
<td>Comparison of survival curves</td>
<td>K Ch 1, 2.1 – 2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>V 3.5.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>K 2.4 – 2.6</td>
</tr>
<tr>
<td>SU – 3</td>
<td>Mon, 4/12</td>
<td>Cox proportional hazards model</td>
<td>V 6.1 – 6.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>K Ch 3</td>
</tr>
<tr>
<td>Wed, 4/14</td>
<td></td>
<td>Review Midterm 2 Cox proportional hazards model</td>
<td></td>
</tr>
<tr>
<td>Mon, 4/19</td>
<td></td>
<td>Homework 7 lab</td>
<td>Homework 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>K Ch 4</td>
</tr>
<tr>
<td>Wed, 4/21</td>
<td></td>
<td>Reading Day No class</td>
<td></td>
</tr>
<tr>
<td>SU – 4</td>
<td>Mon, 4/26</td>
<td>Cox proportional hazards model: Diagnostics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Homework 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>V 6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>K Ch 4</td>
</tr>
<tr>
<td>SU – 5</td>
<td>Wed, 4/28</td>
<td>Power and sample size estimation for the log-rank test</td>
<td></td>
</tr>
<tr>
<td>Mon, 5/3</td>
<td></td>
<td>Homework 8 lab</td>
<td></td>
</tr>
<tr>
<td>Wed, 5/5</td>
<td></td>
<td>Logistic versus Cox versus Poisson Review</td>
<td>Homework 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final Exam Distributed</td>
<td></td>
</tr>
<tr>
<td>Tues, 5/11</td>
<td></td>
<td>Final Examination Due</td>
<td>By Midnight</td>
</tr>
</tbody>
</table>
**Statistical Software:** You will need to use STATA or SAS or R to be able to complete the analyses required for the course. STATA and SAS will be emphasized during the lectures. R code also will be available. Labs will be specific to STATA or SAS/R.

STATA (must be purchased or used in the computer labs)
- Command line based – you put in the *correct* command and get an answer
- Easier to learn
- Appropriate for a wide variety of statistical analyses
- User-added programs are available

SAS (SAS 9.4 must be purchased or used in the computer labs; SAS University is free)
- Used more often in the pharmaceutical industry and in most local and state public health departments
- Data input and manipulation uses a DATA step
- Statistical analysis uses PROC steps
- Steeper learning curve
- Superior data management capabilities than STATA

R (free download)
- Wide variety of statistical and graphical techniques
- Well-designed publication-quality plots
- Widely used for those developing statistical methodology

Select one of the packages for most of your homework and exams and attend the appropriate laboratory section.

A set of modules for learning STATA and SAS are on the D2L site. All students enrolled in the course have access to these modules. Instructions to open the modules are on the D2L site.

A one-unit course “Introduction to Statistical Analysis using STATA” (BIOS 503) is available to those students who view all of the tutorials and complete an online exam for each module. You will need to register for the course to have access to the online exams and to receive credit for the course.

A one-unit course “Introduction to Statistical Analysis using SAS” (BIOS 504) is available to those students who view all of the tutorials and complete an online exam for each module. You will need to register for the course to have access to the online exams and to receive credit for the course.

**Computer Labs:** Stata and SAS are available for public use at two locations:

**Drachman Hall Computing Lab:** Drachman A319, open weekdays, from 8-5. Please bring your own paper for printing.

**Arizona Health Sciences Library Computer Lab:** AHSC 2150, open every day 6:30 am-midnight. These computers are behind the information/reference desk on the main floor. The first couple of banks of machines is not part of the lab, but is rather used for lit searching, etc. The lab is the ‘walled off’ section of computers behind the first couple of banks. You may print output here for a fee. Note that these are public facilities, and may or may not be crowded on a given day.
Students must register to use the AHSC Library Computer Lab at the Library Information Desk. A University of Arizona Catcard is required.

**Purchasing STATA or SAS:**

**STATA:** You can order online at the following link:

http://stata.com/order/new/edu/gradplans/student-pricing

STATA 16 is available. STATA/IC 16 can be purchased at an academic rate of $48 for a 6-month license, $94 for a one-year license and $225 for a perpetual license. After you purchase STATA you will be given directions on downloading and installing the package, with the necessary activation key.

STATA/IC 15 is available in the classroom and lab. It is appropriate for virtually all analyses except for very large datasets with an extremely large number of variables.

PDF versions of the manuals can be accessed from the Help Tab within STATA. There is no need to purchase the manuals.

**SAS:** SAS 9.4 can be ordered from the U of A BookStore. The cost is $99 per year. The link with the necessary information is: http://uabookstore.arizona.edu/technology/stulicense.asp

SAS also has a free “University Edition” of the SAS package. The link with the necessary information is: http://www.sas.com/en_us/software/university-edition.html

SAS also has a free “SAS OnDemand for Academics” version of the SAS package. It is a cloud-based program so you do not need to download anything to your computer. However, since it is cloud-based it may be slower during times of high demand. Here are the steps for using it:

1. As a first step, please create your account for SAS OnDemand for Academics. To register, visit https://odamid.oda.sas.com and click on Register for an account.
3. Look for the Enroll in a course link in the ”Enrollments” section near the bottom of the page. Click this link to start the enrollment.
4. Enter the course code: f45fa61d-4c40-4487-8e8c-ac0d6b9885bd
5. Submit the form.
6. Confirm that this is the correct course and then click the button to finish enrolling.

Students who plan to use SAS in the workplace should learn SAS 9.4 as the “University Edition” is not available outside the University.

The current notes are based on the "full version" of SAS. Please alert me if there are substantial differences between the notes and the “University Edition”.

**R:** You can download R at the following link:

https://www.r-project.org/
Tips for Succeeding in the Course:

1. Attend class or review the lecture online
2. Ask questions about the lectures and notes
3. Do your homework early so that you can ask questions when it is reviewed in the lab session
4. Turn your homework, Midterms and Final Exam in on time
5. Ask questions until you understand the material

Classroom Behavior Policy: (In-person classes only) To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Communications: You are responsible for reading emails sent to your UA account from your professor and the announcements that are placed on the course web site. Information about readings, news, your grades, assignments and other course related topics will be communicated to you with these electronic methods. The official policy can be found at: https://www.registrar.arizona.edu/personal-information/official-student-email-policy-use-email-official-correspondence-students

UA Smoking and Tobacco Policy: The purpose of this Policy is to establish the University of Arizona’s (University) commitment to protect the health of University faculty, staff, students, and visitors on its campuses and in its vehicle. The official policy can be found at: http://policy.arizona.edu/ethics-and-conduct/smoking-and-tobacco-policy

University-wide Policies link: Links to the following UA policies are provided here: https://academicaffairs.arizona.edu/syllabus-policies

• Absence and Class Participation Policies
• Threatening Behavior Policy
• Accessibility and Accommodations Policy
• Code of Academic Integrity
• Nondiscrimination and Anti-Harassment Policy
• Subject to Change Statement

Class Attendance/Participation: Class attendance (via Zoom/Panopto or in person) and/or reviewing the recorded lectures is strongly encouraged, but not required. If a student misses class, they are responsible for meeting all course deadlines, and for working with other students, the TA and the instructor (during office hours) to catch up.

Academic Integrity: All students are expected to do their own work. For homework, feel free to ask each other questions about concepts and procedures. However, when it comes time to complete the homework to turn in, do that on your own. Duplicate homework will be considered a breach of academic integrity. No communication between students of any sort is allowed during exams or the final.
**Plagiarism:** What counts as plagiarism?
- Copying and pasting information from a web site or another source, and then revising it so that it sounds like your original idea (beware of Wikipedia).
- Doing a homework assignment with a friend and then handing in separate assignments that contain the same ideas, language, phrases, etc.
- Paraphrasing a passage without citing it, so that it looks like your own.
- Hiring another person to do your work for you, or purchasing a paper through any of the on-or off-line sources.

**Syllabus Changes:** Information contained in the course syllabus, other than the grade and absence policies, may be subject to change with reasonable advance notice, as deemed appropriate by the instructor.