BROWN SCHOOL

**WASHINGTON UNIVERSITY IN ST. LOUIS**

**MASTER OF PUBLIC HEALTH PROGRAM**

**SPRING 2018**

**Special topics: Advanced data analysis**

**CREDIT HOURS:** 2 **INSTRUCTOR:** Kimberly J. Johnson, MPH, PhD

**GRADE:** L/G **OFFICE:** Goldfarb 237 (Instructor)

**ROOM:** Goldfarb 330 **OFFICE HOURS:**  Kim: Wednesdays 4-5 pm G237 or by appointment

Xiaoyan: 11-12 pm Rm 104 of Brown Hall

**TIMES:** Thursday 1-3 **PHONE (Instructor):** 314-935-9154 (office), 314-809-9452 (cell)

**TA:** Xiaoyan Wang **E-MAILS:** [kijohnson@wustl.edu](mailto:kijohnson@wustl.edu)

[xiaoyan.wang@wustl.edu](mailto:xiaoyan.wang@wustl.edu)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_­­­­­

1. **COURSE DOMAIN AND BOUNDARIES**

Students will be provided with an introduction to theory and analytical techniques commonly employed in the practice of public health through mini-lectures and a hands-on experiential learning approach. The flow of the course will follow that of the format frequently used in the production of public health reports starting with preparing your data for analyses and ending with communicating your results. Students will work as individuals or teams to develop and conduct a project, choosing from different real-world public health datasets that are amenable to some of the most common types of analyses that students will encounter in their public health careers. Mini-lectures will address theory as well as practical tools and in class exercises will be focused on application of the theories and tools. Homework will be assigned that provides students with further opportunity to gain competency in strategies that can be employed to analyze each type of dataset. The course will also provide students with the opportunity to skills in data visualization strategies, conducting reproducible research, presenting projects, and effectively responding to constructive critiques.

Prerequisites:

- Foundations of Public Health: Epidemiology

- Foundations of Public Health: Biostatistics

- Applied Linear Modeling

- Basic competency in R or motivation to learn R in those with competency in SAS, SPSS, or STATA

**II. CORE COMPETENCIES AND PRACTICE BEHAVIORS**

**III. BROWN SCHOOL ACADEMIC POLICIES**

Academic Integrity: Academic integrity in the completion of tests, oral presentations and written assignments (including statistical syntax) is expected. Violations of academic integrity (e.g., plagiarism) are very serious offenses. Violations will result in notification to the Assistant Dean for the appropriate (MSW or MPH) program and may result in referral to the Academic and Professional Integrity Committee, which could lead to dismissal from the program. Please review and adhere to the entire set of Academic Integrity guidelines in the student handbook on Inside Brown:

[Student Handbook 2017-2018](https://insidebrown.gwb.wustl.edu/People/students/studenthandbook/Pages/default.aspx)

Accommodations: If you have a learning disability, sensory, or physical disability or other impairment, and you may need special assistance in lectures, reading, written assignments, and/or exam taking, please contact the Brown School Director of Student Affairs who can provide coordination of accommodations at Washington University and the Brown School. The [Disability Resource Center](http://cornerstone.wustl.edu/DisabilityResources.aspx), a University-wide resource, provides diagnostic and academic accommodations support and referrals.

English Language Proficiency: If your English language proficiency is such that you may need special assistance in lectures, reading, written assignments, and/or exam taking, please communicate these needs to your instructor who may refer you to the [English Language Program](http://oiss.wustl.edu/english-language-programs/) (ELP), a University-wide resource which provides classes and academic English language support designed to increase non-native English speaking students' English language proficiency and to facilitate their academic success at Washington University. You may also find the Academic Assistance resources available through the [Office for International Students and Scholars](http://oisshome.wustl.edu/students/) to be helpful.

Professional Use of Electronic Devices in the Classroom: Computers or other electronic devices, including “smart pens” (devices with an embedded computer and digital audio recorder which records the classroom lecture/discussion and links that recording to the notes taken by the student), may be used by students at the discretion of the faculty member to support the learning activities in the classroom. These include such activities as taking notes and accessing course readings under discussion. If a student wishes to use a smart-pen or other electronic device to audio record lectures or class discussions, they must notify the instructor in advance of doing so. Permission to use recording devices will be at the discretion of the instructor, unless this is an accommodation approved by Disability Resources.

Nonacademic use of laptops and other devices is distracting and seriously disrupts the learning process for everyone. Neither computers nor other electronic devices are to be used in the classroom during class for non-academic reasons. This use includes emailing, texting, social networking, and use of the Internet. The use of cell phones during class time is prohibited, and they should be set on silent before class begins. In the case of an emergency, please step out of the room to take the call. The instructor has the right to hold students accountable for meeting these expectations, and failure to do so may result in a loss of participation points, a loss of the privilege of computer use in the classroom, or being asked to leave the classroom.

Religious Holidays: The Brown School recognizes the individual student’s choice in observing religious holidays that occur during periods when classes are scheduled. Students are encouraged to arrange with their instructors to make up work missed as a result of religious observance, and instructors are asked to make every reasonable effort to accommodate such requests.

**IV. WASHINGTON UNIVERSITY ACADEMIC SUPPORT POLICIES**

Accommodations based upon sexual assault: The University is committed to offering reasonable academic accommodations to students who are victims of sexual assault.  Students are eligible for accommodation regardless of whether they seek criminal or disciplinary action.  Depending on the specific nature of the allegation, such measures may include but are not limited to: implementation of a no-contact order, course/classroom assignment changes, and other academic support services and accommodations.  If you need to request such accommodations, please direct your request to Kim Webb ([kim\_webb@wustl.edu](mailto:kim_webb@wustl.edu)), Director of the [Relationship and Sexual Violence Prevention Center](http://mailingsresponse.wustl.edu/trk/click?ref=z1030up2e7_1-911fx36678x01649&).  Ms. Webb is a confidential resource; however, requests for accommodations will be shared with the appropriate University administration and faculty.  The University will maintain as confidential any accommodations or protective measures provided to an individual student so long as it does not impair the ability to provide such measures.

If a student comes to me to discuss or disclose an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if I otherwise observe or become aware of such an allegation, I will keep the information as private as I can, but as a faculty member of Washington University, I am required to immediately report it to my Department Chair or Dean or directly to Ms. Jessica Kennedy, the University’s Title IX Director.  If you would like to speak with directly Ms. Kennedy directly, she can be reached at [(314) 935-3118](tel:(314)%20935-3118), [jwkennedy@wustl.edu](mailto:jwkennedy@wustl.edu), or by visiting the [Title IX office](http://mailingsresponse.wustl.edu/trk/click?ref=z1030up2e7_1-911fx36679x01649&) in Umrath Hall.  Additionally, you can report incidents or complaints to the Office of Student Conduct and Community Standards or by contacting WUPD at [(314) 935-5555](tel:(314)%20935-5555) or your local law enforcement agency. See: [Title IX](http://mailingsresponse.wustl.edu/trk/click?ref=z1030up2e7_1-911fx3667ax01649&)

You can also speak confidentially and learn more about available resources at the Relationship and Sexual Violence Prevention Center by calling [(314) 935-8761](tel:(314)%20935-8761) or visiting the 4th floor of Seigle Hall.  See: [RSVP Center](http://mailingsresponse.wustl.edu/trk/click?ref=z1030up2e7_1-911fx3667bx01649&)

Bias Reporting: The University has a process through which students, faculty, staff and community members who have experienced or witnessed incidents of bias, prejudice or discrimination against a student can report their experiences to the University’s Bias Report and Support System (BRSS) team.  See:  [brss.wustl.edu](http://mailingsresponse.wustl.edu/trk/click?ref=z1030up2e7_1-911fx3667cx01649&)

Mental Health: Mental Health Services’ professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety and depression. See [shs.wustl.edu/MentalHealth](http://mailingsresponse.wustl.edu/trk/click?ref=z1030up2e7_1-911fx3667dx01649&)

Additional Issues or Concerns: If you feel that you need additional supports in order to be successful in your time at Brown, beyond the afore mentioned accommodations, please contact Essie Rochman, Director of Student Affairs at [erochman@wustl.edu](mailto:erochman@wustl.edu). She can assist you in navigating a myriad of concerns. Her office is in Brown Hall, room 320.

1. **READINGS AND SOFTWARE REQUIREMENTS**

**Readings:** The reading assignments for the course will include links to journal articles and other material provided on our class Github website: <https://github.com/kijohnson/Advanced-Data-Analysis>.

**Software:** Students are required to use R for this course for analyses. Instructions for how to gain access is provided on the Github website: <https://github.com/kijohnson/Advanced-Data-Analysis>.

1. **ORGANIZATION OF THE COURSE**

The format of the course will be lectures, examples, and applied in class activities, and discussion. You will perform exercises that reinforce the epidemiologic and biostatistics principles and methods that are presented in class. Your mastery of the course objectives will be assessed through quizzes, homework assignments and a final course project.

1. **ROLE OF STUDENT, INSTRUCTOR, and TAs**

Student Expectations and Requirements**:**

* You are expected to read any assigned materials prior to class.
* You should be prepared to discuss the assigned readings at the start of each class and to participate in class discussions.
* You must return your assignments on time.
* You are responsible for doing your own work.
* You are expected to check blackboard and email regularly for any course announcements.
* If you have any issues or challenges with the class, please come to see me or the TA so that we can resolve them as quickly as possible*.*

Instructor Expectations and Requirements:

* I will come to class prepared, organized, and enthusiastic.
* I can be reached by phone, email, or during office hours/scheduled appointments.
* I retain the right to change the content and order of the lectures and exercises to meet the needs of students who are enrolled in the course.

Role of TA:

* The TA will assist the instructor with preparation of materials for class, grading student assignements/exams, and will hold office hours weekly. The TA(s) will also be available by email to answer questions or facilitate answering of questions.

1. **ASSIGNMENTS AND GRADING CRITERIA**

**Homework (25% of final grade).** There will be 12 homework assignments (most will be relatively short). Students may drop 2 of the 12 assignments.

**Midterm Exam (15% of final grade).** This exam will be conducted online during class to assess student learning up to week 8.

**Statistical code submission and peer review (10% of final grade):** Submit well-formatted and annotated statistical code and run and review the statistical code of two classmates (5 total peer review opportunities, four will count as 2.5 points each).

**Final project (40% of final grade):** Students will develop a data analysis project from conception to dissemination of results through a presentation and report using a real-world dataset. Students will find their own datasets. Examples of real-world datasets that could be used for this project included Surveillance, Epidemiology, and End Results (SEER: <https://seer.cancer.gov/>), Behavioral Risk Factor Surveillance System (BRFSS: <https://www.cdc.gov/brfss/index.html>), National Health and Nutrition Examination Survey (NHANES: <https://www.cdc.gov/nchs/nhanes/index.htm>). Students are expected to apply the concepts used in class to their final project.

The final project can be conducted by an individual or a team of no more than 3 members. For students choosing to work in teams, each team member will receive the same grade. It is therefore advised that you choose your team members with careful consideration as I will not take off points for individual team members who may cause other team members frustration! ☺ Components of the project include:

1. **Concept proposal (5 points)**
2. **Abstract (5 points)**
3. **Code for project on github site (5 points)**
4. **Final project presentation (10 points)**
5. **Final project report (10 points)**
6. **Responses to reviewers (5 points).** Reviewers will be composed of students in the class and/or 2-3 faculty/staff members who will provide written comments that the instructor will assemble and provide to the student/team. Responses are due within one week of the final presentation.

**Participation and professionalism (10% of final grade).** Participation will be assessed based on **attendance.** For full credit, students are required to attend 14/15 classes to receive full attendance points. Attendance will be tracked. Permission to miss additional classes may be granted without the subsequent loss of points. Professionalism is considered in the final participation grade. Professionalism considerations include effective participation in team work and with the instructor to resolve learning issues, timely responses to email, participation in discussions, and attentiveness in class.

**Extra Credit:** ***Very nomimal extra credit*** (0.1 points toward final grade) but ***very real bragging rights*** will be awarded during each peer review exercise to the person who writes the most efficient piece of code and to the person who writes the most inefficient piece of code *that still works*. If you are a super R user already, you may want to challenge yourself and try writing some very inefficient code. Efficiency of code will measured as the fewest steps and characters for code to run while still following good coding practices. Time to run may also be factored in. In addition to the very nominal extra credit and the very real bragging rights, winners will each get a chocolate medallion.

**Grading Scale:**

95-100: A\* 90-94: A- 88-89: B+ 82-87: B

80-81: B- 77-79: C+ 74-76, C 70-73: C-

69 & below, F

**\*What constitutes A work:** As indicated in “Course Administration and Grading” (<https://insidebrown.gwb.wustl.edu/students/MPH/Pages/AcademicResources.aspx> ), A’s are given for outstanding work defined as “**exceeds expected level of performance for graduate study”.** Not all students will receive A’s in this course and extra credit assignments are not given to assist students in recovering lost points. In my experience students who perform at the outstanding level in this class are typically those that come to class prepared, start their homework well before the due date, and ask questions to clarify gaps in their understanding. If you are not performing at an A level of work, please know that I do not use grades to judge the overall quality of the person or as markers to predict that person’s success. There are many factors besides grades that contribute to success and I do realize that students have varying priorities and demands on their time. I view each student wholistically and grades are just one part of the overall view.

1. **Getting help with R and course content in general.**

If you need help with understanding the course materials, there are several options including:

* Making an appointment with me or the TA.
* Making an appointment with the StatLab that provides excellent tutoring with software and programming. Please visit the StatLab website for more information at <https://insidebrown.gwb.wustl.edu/People/students/StudentSupportResources/Pages/StatLab.aspx>
* Good websites for R issues:
  + Quick-R: <https://www.statmethods.net/> . I often consult this website for basic R coding how-to’s.
  + UCLA Institute for Digital Research and Education (idre): <https://stats.idre.ucla.edu/r/> . I haven’t used this much for R but I have used it for SAS and have found it to be very helpful for training me how to do something in SAS. For example, I learned how to do survival analysis diagnostics on this site as well as multiple imputation methods.
  + Stackoverflow: <https://stackoverflow.com/> (please see: <https://stackoverflow.com/tour> before posting a question on this site. Users of this site get very crabby when questions are not well-written (clearly telling what problem you are trying to solve and showing what you already tried) so make sure you read about what not to ask about at the tour link.
  + Google your error message or your “how do you…” question and often you will find a solution on the web. This is a completely normal part of coding so embrace it!

1. **COURSE OUTLINE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class: Date** | **Readings** | **Topic (s)** | **Lecture** | **In class exercises** | **What is due?** |
| **1:** 1/18 | * None | Course overview (e.g., syllabus, Blackboard), introduction to fellow students,  Getting started | The project life cycle/R review (Kim Johnson, MPH, PhD) | Analysis of survey data | **HW1:** Survey responses and installation of R and R studio |
| **2:** 1/25 | * Peng RD. Reproducible Research in Computational Science. *Science* Vol 334, 2011. Pg. 1226-1227 * Knuth DE. Literate Programming. [*The Computer Journal*](https://en.wikipedia.org/wiki/The_Computer_Journal).  (read pg 1 part A) * Analysis and Reporting Modules 3c and 3d   [3c](https://www.youtube.com/watch?v=-_oNryFHh8c&list=PLkfBg8ML-gIm4PsDBIplBONOyMon2p9k-&index=7)  [3d](https://www.youtube.com/watch?v=qaRW2-Dy-F0&list=PLkfBg8ML-gIm4PsDBIplBONOyMon2p9k-&index=8) | Reproducible research I | Reproducible research (Kim Johnson, MPH, PhD) | Analyzing data using R markdown | **HW2:** Survey analysis questions |
| **3:** 2/1 | * Ram K. Git can facilitate greater reproducibility and increased transparency in science. Ram Source Code for Biology and Medicine 2013, 8:7 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3639880/> * Peng videos-these should really have gone last week but I wanted to spread out the content!   Peng 1: <https://www.youtube.com/watch?v=4rBX6r5emgQ>  Peng 2: <https://www.youtube.com/watch?v=VOaN33aAcEw>  Peng 3: <https://www.youtube.com/watch?v=VOaN33aAcEw> | Reproducible research II | Using git for version control (Kim Johnson, MPH, PhD) | Installing and using Git | **HW3:** Reproducible research assignment (R markdown and code annotation of height assignment)  Peer review 1 |
| **4:** 2/8 | * The visual display of quantitative information / Edward R. Tufte. Edition. 2nd ed. Published Cheshire, Conn. : Graphics Press, c2001. *Read Chapter 1* | Data Visualization | Data visualization in R for scientific presentations (Matt Wyczalkowski, PhD) | Data visualization exercise | **HW4:** Git exercise |
| **5:** 2/15 | * Quick-R Generalized Linear Models <https://www.statmethods.net/advstats/glm.html> | Generalized linear models | Overview of generalized linear models (Joe Steensma, EdD) | Generalized linear models in R exercise | **HW5:** Data visualization assignment |
| **6:** 2/22 | **TBD** | Generalized linear modeling II | Multinomial and ordinal logistic regression (Xiaoyan Wang) | Multinomial and ordinal logistic regression excercise | **HW6:** Generalized linear models assignment  Peer review 2 |
| **7:** 3/1 | **TBD** | Generalized linear modeling III | Poisson and negative binomial regression (Shenyang Guo, PhD) | Poisson and negative binomial regression exercise | **HW7:** Multinomial and ordinal logistic regression assignment  Concept proposal for project |
| **8: 3/8** | **Midterm exam** | | | | |
| **9:** 3/22 | * A Practical Guide to Understanding Kaplan-Meier Curves <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3932959/pdf/nihms549224.pdf> | Survival analysis I | Introduction to Survival Analysis (Kim Johnson, MPH, PhD) | Kaplan Meier Curves in R exercise | **HW8:** Poisson and negative binomial regression assignment  Peer review 3 |
| **10:** 3/29 | **TBD** | Survival analysis II | Cox Proportional Hazards Regression (Shenyang Guo, PhD) | Cox PH in R exercise | **HW9:** KM curves  Peer review 4 |
| **11:** 4/5 | * Suttorp MM et al. Graphical presentation of confounding in directed acyclic graphs. Nephrol Dial Transplant. 2015 Sep;30(9):1418-23. doi: 10.1093/ndt/gfu325. Epub 2014 Oct 16. (REQUIRED) * Westreich D and Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am J. Epidemiol 2013 Feb 15;177(4):292-8. PMID: 23371353 (OPTIONAL) * VanderWeele TJ and Robinson W. On causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology. 2014 July; 25(4): 473-484 (OPTIONAL) * Shrier I and Platt RW. Reducing bias through directed acyclic graphs. BMC Medical Research Methodology. 2008, 8:70 (OPTIONAL) | Covariate selection (DAGs) | DAGs (Kim Johnson, MPH, PhD) | DAG exercise | **HW10:** Cox PH assignment |
| **12:** 4/12 | * Cummings P. Missing Data and Multiple Imputation JAMA Pediatrics. July 2013 Volume 167, Number 7 656-661. | Other critical issues | Sensitivity analyses/Missing DATA (Kim Johnson, MPH, PhD) | Sensitivity analyses for selection bias and missing data | **HW11:** DAG assignment |
| **13:** 4/19 | * Nuzzo R. Scientific method: statistical errors. Nature 2014 Feb 13; 506 (7487) * <http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108?needAccess=true> * Jan P. Vandenbroucke, Erik von Elm, Douglas G. Altman, Peter C. Gøtzsche, Cynthia D. Mulrow, Stuart J. Pocock, Charles Poole, James J. Schlesselman, and Matthias Egger, for the STROBE Initiative, “Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration,” 18 Epidemiology 805, 817 (2007). | Good science | P-values and reporting (Kim Johnson, MPH, PhD) | Working class period | **HW12:** Missing data assignment  Peer review 5 |
| **14:** 4/26 | Final project presentations with ‘shark tank’ panel review |  |  |  |  |
| **15:** 5/3 | Final project presentations with ‘shark tank’ panel review |  |  |  |  |
| **5/10** |  |  |  |  | Final project revisions |