

College of Health and Human Services

**FALL 2023**

|  |  |
| --- | --- |
| **Syllabus**  **HAP 819: Advanced Statistics in Health Services Research II** | |
| **Course information** | **Day/time:** Hybrid (in person and online sessions via Zoom):  **Location:** In-person sessions, Peterson Hall 2500; See Zoom location in the meeting invite |
| **Instructor** | **Name:** Farrokh Alemi, PhD  **Contact info:** E-mail: falemi@gmu.edu, (7038933799, text first before calling)  **Office location:** Peterson Hall  **Office hours:** By appointment (via phone, in-person or Zoom)  **Teaching Assistant:** TBD |
| **Course placement** | **(X)** Core **( )** Concentration **()** Elective **( )** Pre-requisite(s)  **(X )** *Course(s) recommended before taking this course*: *HAP 719* |
| **Course description** | *Prerequisite: HAP 719 or equivalent graduate statistics course.*  This course covers principles and methods of advanced statistical data analysis and inference with applications in health services research. It emphasizes the use and application of various data analysis techniques, including multivariate statistics, regression and longitudinal data analysis. |
| **Course objectives** | 1. Explain multivariate relationships and their application in data analysis. 2. Determine when to use repeated measures analysis. 3. Perform multivariate analysis of variance and covariance. 4. Determine when a specific type of regression should be used based on the level of the outcome variable. 5. Examine situations in which longitudinal data analysis should be used and perform data analyses related to this methodology. 6. Examine how time varying exposure affects analysis of outcomes of care 7. Examine situation in which LASSO regressions should be used and perform data analyses using it 8. Examine situations in which propensity weighted regression should be used and perform data analysis using it 9. Examine situations in which multi-level modeling should be used and perform data analysis using it. 10. Examine situations in which covariate balancing should be used in analysis of observational data 11. Examine missing value imputation and stratified missing values 12. Interpret results from statistical analysis 13. Foster writing reports and communication of results of data analysis. |
| **textbook and/or materials** | 1. Required: Alemi F. Big Data in Healthcare: Statistical Analysis of the Electronic Health Record (1) 1st Edition. Available at <https://www.amazon.com/Big-Data-Healthcare-Statistical-Electronic/dp/1640550631> 2. Recommended: Huber M. Causal Analysis: Impact Evaluation and Causal Machine Learning with Applications in R. The MIT Press, 2023. 3. Course open site is at [http://openonlinecourses.com/819 after 8/30/2023](http://openonlinecourses.com/819%20after%208/30/2023) |
| **Statistical software** | 1. You may use the STATA User’s Guide for Version 16: <https://www.stata.com/manuals/u.pdf>. Students may purchase a copy of STATA at <http://www.stata.com/order/new/edu/gradplans/student-pricing/> 2. You may use freely available R software. Examples in the course are based on R. We provide support for R only but students can use any software. |
| **Data set** | To demonstrate the use of several statistical methodologies and develop data analysis skills using a statistical software, we will use data from All of Use database. You must register for this database within first week of class. Details on how to register are provided at <https://researchallofus.org/> |
| **Teaching methods** | This course meets on announced days for 3 hours. The first hour is a lecture on the topic. The second hour is a lab, where students work together to do assignments. The third hour is set aside for students who need to teach next week. They work ahead and prepare their solutions for assignments in following week.  Students are used to seeing data cleaned by their instructor, usually toy data that is small in size. In this course we provide you with large massive data that is not clean and that you need to spend time to clean it. The processing of these data also takes time. So many students complain to me about it. I hear what you are saying. You want to learn statistics and not data cleaning or processing of massive data. Bear with me. This course is different in trying to push you to feel comfortable in analysis of massive data. It is also different for spending significant time in data preparation. Why?  Because real statistics is 80% of time about preparation of data and the final modeling step is a minor step. I want you to feel what it takes to do all of the steps in the data analysis. If you focus on data modeling step, usually a single line, you would miss the real place where data is alive and changing. It is a false sense of comfort when you learn data modeling in small, previously cleaned, data. You miss the real issues of how to fix missing values, how to transform data to meet assumptions of the model, how to take out unnecessary data. These are not trivial issues, they are at the core of understanding. To learn modeling concepts, you still need to transform data to get it ready.  I call this torturing data until they confess. I know it is time consuming to learn the concepts and related data preparations all at the same time but there is no better way. Real statisticians do not separate data modeling from data transformation and preparation.  We often share screens. Mac Users often have difficulty sharing R studio screen. To do so, Search "recording" on System Preferences - allow your browser (e.g., Chrome) to do screen recording. Restart browser. |
| **Evaluation &**  **course requirements** | |  |  | | --- | --- | | It is expected that each of you will:  1. Read all assigned materials  2. Complete the assignments  3. Complete a major project  4. Complete a teach-one presentation  5. Complete an exam  6. Attend each class session  7. Must activate & use GMU e-mail | Late assignments are not accepted by the instructor. If an emergency occurs, please notify the instructor in advance. | |
| **Grading scale** | |  |  | | --- | --- | | **Graded Activities** | **Grade points** | | Weekly assignments | 10 pts each/40 pts | | Semester long project | 25 pts | | Teach One presentation | 15 pts | | Exam (may be waived) | 20 pts | | Total | 100 pts or 100% | | There is no curving of grades | | |
| **Grade assignment** | |  | | --- | | A+ = 98-100, A = 93-97, A- = 90-92 | | B+ = 87-89, B = 83-86, B- = 80-82 | | C+ = 77-79, C = 73-76, C- = 70-72 | | D = 60-69, F = 0-59 | |
| **assignments (4)** | A set of homework problems will be assigned to evaluate your skills in applying concepts learned during lectures. You can work with others in doing these assignments but submissions must be entirely your code, no copy and pasting. Assignments are due on following class session. No late assignments are accepted to make sure that the entire class stays focused on same set of assignments. |
| **group project (1)** | The project is designed to provide a review of concepts learned throughout the semester and exercise student ability in applying these concepts in analyzing data sets, writing up results from the statistical analyses, and interpreting findings. This is a data analysis project. The length of the project report is expected to be similar to a journal submission, you choose the journal you would like to use as a template. This is an individual project. The written document should include:   1. Background literature review should not exceed 1 page. It should assume an informed reader familiar with the literature. It should reference all relevant recently published material on this topic in PubMed. Background section should be a brief synthesis of existing research findings related to the problem being addressed in the study. This section is due in 3rd week of the course. 2. Method section should be a complete description of the methods, including: source of data, sample description, standardized measures used, analytical methods used, and adjustments for missing values. This section is due on 5th week of the course. 3. Results section should describe the findings. Table 1 should be description of the population studied. Figures and additional tables should be the statistical findings. There should not be any description of findings in the result section. This section should be complete 10th week of the course. 4. Discussion section should in 4-5 paragraphs provide: (1) summary of the key findings, (2) describe whether the literature supports these findings, (3) provide study limitations, and (4) indicate policy implications. |
| **Teach One presentation**  **(1)** | * Students should work with the instructor one week prior to class to present a recorded and narrated slides on one of the topics in the class. * Teach-one presentations are graded based on peer-groups ranking of how helpful the student-teacher was. |
| **Exam (1)** | * One final take-home exam will be given at the end of the semester.This is a cumulative exam covering all topics from the beginning of the semester. Students are allowed to use the textbook, lecture notes and homework in answering exam questions. However, no other help (e.g., classmates, tutors) is allowed. If you wish to cancel the exam, you need to show that on the project, weekly assignments and teach one assignment you are receiving a grade of A. |
| **Attendance & course participation** | * Attendance will not be taken in this course. Because of the importance of lecture and discussion to your total learning experience, I wish to encourage you to both attend and participate in class regularly. Attendance, punctuality, preparation, and active contribution to small and large group efforts are essential. * Students, who miss an assignment must notify the instructor (preferably in advance) and are responsible for completing all assignments and readings for the next class. * Late assignments will not be accepted unless a serious emergency arises and the instructor is notified promptly. |
| **Mason Honor Code** | To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University Community and with the desire for greater academic and personal achievement, we, the student members of the university community, have set for this Honor Code: Student Members of the George Mason University community pledge not to cheat, plagiarize, steal, or lie in matters related to academic work.A full reading of the Honor Code and the associated system can be found at our website, [*oai.gmu.edu*](http://oai.gmu.edu/). |
| **Individuals with Disabilities** | The Office of Disability Services (ODS) collaborates with students with documented disabilities and faculty to provide reasonable accommodations, auxiliary aids, and support services that are individualized and based upon medical documentation, functional limitations, and a collaborative assessment of needs. In order to receive accommodations, students must complete the following process: **https://ds.gmu.edu/eligibility/** |
| **E-mail policy** | * Web: <https://its.gmu.edu/service/masonlive-email-for-students/>. Mason uses electronic mail to provide official information to students. Examples include notices from the library, notices about academic standing, financial aid information, class materials, assignments, questions, and instructor feedback. Students are responsible for the content of university communication sent to their Mason e-mail account and are required to activate that account and check it regularly. Students are also expected to maintain an active and accurate mailing address in order to receive communications sent through the United States Postal Service. * Instructor is available on text between 9 am to 9 pm at 703 893 3799. On day 1, you must send a text so that you are recognized when you text in the future. |
| **Return-to-Campus Policies** | All students taking courses with a face-to-face component are required to follow the university’s public health and safety precautions and procedures outlined on the university Safe Return to Campus webpage (<https://www2.gmu.edu/safe-return-campus>). Your instructor has prepared an AI system for checking your COVID symptoms: <https://autonosis.io/>. Information about Mask Policy, Covid Health Check, Covid testing: <https://www.gmu.edu/safe-return-campus/mason-covid-updates/updates-students> |
| **Other useful campus recourses** | * University Policies: <http://universitypolicy.gmu.edu/> * Writing Center: http://writingcenter.gmu.edu * Digital scholarship Center: <https://dsc.gmu.edu/help/> * Health Sciences Librarian: Kathy Butler Science & Technology Team: George Mason University [kbutle18@gmu.edu](mailto:kbutle18@gmu.edu) * University Libraries: <https://www2.gmu.edu/academics/university-libraries> * Counseling and Psychological Services: (703) 993-2380; <http://caps.gmu.edu> * **Use of ChatGPT and other AI systems is encouraged** |



**Project Rubric**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criteria** | **0%**  **Unacceptable** | **50%**  **Needs Improvement** | **75%**  **Competent** | **100%**  **Exemplary** |
| Problem, scope, and/or hypothesis are clearly stated. Project includes a full explanation of topic and background. | No explanation of problem, background, or aspects of the topic.  Scope and/or hypothesis were not stated. (0 points) | Problem/issue present, but unclear.  Background information omitted; incomplete explanation of topic.  Few aspects/limitations identified or explored.  Scope and/or hypotheses were not articulated well. (10 points) | Problem adequately identified and described.  Adequate explanation of topic; some important background information omitted.  Most aspects/limitations/gaps considered.  Scope and/or hypothesis were stated. (15 points) | Problem clearly identified. Full explanation of topic and background via literature review.  All aspects and limitations/gaps in current research (based on literature review) considered. Scope of project and/or hypotheses were clearly stated. (20 points) |
| Provides a detailed description of the data set and design including; sample size information; identifies and describes dependent and independent variables and statistical methodology to analyze data. | Poor. Major information omitted. Inadequate description of methodology if not omitted. (0 points) | Somewhere between adequate and fair description of data set, design and identification of dependent and independent variables as well as description of statistical methodology. (10 points) | Adequate description of data set and design, and identification of dependent and independent variables as well as description of statistical methodology. (15 points) | Provided a detailed description of the data set and design, sample size, identified and described dependent and independent variables. Described statistical methodology to analyze data. (20 points) |
| Utilizes appropriate statistical techniques including descriptive and inferential statistics to analyze the data. Provides a detailed and correct description of results. Presents results appropriately and clearly in tables and/or graphs. | Poor. Used wrong methodology and poor description of results. (0 points) | Somewhere between adequate to fair use of methodology and description of results. (10 points) | Used appropriate statistical techniques including descriptive and inferential statistics to analyze the data.  Adequate description of results. (15 points) | Used appropriate statistical techniques including descriptive and inferential statistics to analyze the data. Provided a detailed and correct description of results. Results are presented in tables and/or graphs appropriately and in a clear manner. (20 points) |
| Summarizes and synthesizes all points. Conclusions are based on evidence/sound methods. Limitations and implications are addressed. | No synthesis or summary. Conclusions ignored; maintains preconceived views regardless of evidence. Limitations were not addressed. (0 points) | Summary points briefly mentioned, weak synthesis.  Conclusions drawn but errors presented. Limitations were not addressed fully. (10 points) | Some points summarized and synthesized. Most conclusions based on evidence used with adequate methods. Most limitations were addressed. (15 points) | All points summarized and synthesized. Conclusions based on evidence/sound methods. Limitations and implications were addressed. (20 points) |
| Utilizes proper grammar, mechanics, and specific format style throughout. | Improper grammar and mechanics throughout. Improper use of a specific forma style throughout. (0 points) | Weak use of proper grammar and mechanics throughout. Weak use of specific format style throughout. (10 points) | Mostly used proper grammar and mechanics throughout. Adequate use of specific format style throughout. (15 points) | Use of proper grammar and mechanics. Proper use of specific format style throughout. (20 points) |