



University of Maryland School of Public Health

EPIB 661 – Applied Multivariate Data Analysis

Semester: Spring 2019
Section: 0101
Classroom and Time: Monday 1:00-3:45pm, ATL 1114
Course webpage: <https://umd.instructure.com/courses/1261668>
Instructor: Charles Ma, Ph.D. **Office Hours:** Monday 11:00-12:00pm
or by appointment
Office: 2234M
Phone: 301-405-6421
Email: tma0929@umd.edu

Course Description: Multivariate analysis targets at data with simultaneous measurements on many variables and studies the relationship between these variables. This course introduces important multivariate analysis methods used in public health research. Students will learn how to perform multivariate statistical analyses and understand the results. Topics covered include multivariate regression analysis, multivariate analysis of variance (MANOVA), principal component analysis (PCA), factor analysis, classification, clustering analysis, canonical correlation analysis (CCA) and correspondence analysis (CA). For each topic, real examples and programming codes are provided and the emphasis is on practical applications of multivariate analysis methods in the health sciences.

Course Prerequisites:

Required: EPIB 650 *Biostatistics I* and EPIB 651 *Biostatistics II* or permission of instructor.

Recommended: Previous experience with at least one statistical software (e.g. SAS*, R, STATA)

*: SAS is the main software used for demonstration in class

Course Learning Objectives:

Upon completing this course, the student will be able to:

1. Understand the basic concepts of a number of multivariate analysis methods including multivariate regression, MANOVA, PCA, factor analysis, classification, clustering analysis, CCA and CA.
2. Choose the appropriate methods and models for a variety of multivariate analysis problems.
3. Use statistical software to conduct an appropriate multivariate data analysis.
4. Prepare graphical output for PCA, factor analysis, classification, clustering analysis, CCA and CA.
5. Interpret the multivariate analysis results.
6. Make inference about the relationship between one or more response variables and multiple explanatory variables.
7. Check assumptions and perform model diagnostics of the underlying models.

Program Competencies Addressed in this Course:

The following competencies for the *Master of Public Health with concentration in Biostatistics* are addressed in this course:

1. Distinguish among the different measurement scales or types of variables and select appropriate descriptive statistical methods for summarizing public health data.
2. Select appropriate inferential statistical methods to answer research questions relevant to public health research.
3. Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research.
4. Critically evaluate statistical analyses presented in public health literature.
5. Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures.
6. Communicate results of statistical analyses to lay and professional audiences.
7. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate.
8. Interpret results of data analysis for public health research, policy or practice.

Recommended Texts and Other Readings:

Main reference books:

Abdelmonem A. Afifi, Susanne May & Virginia Clark (2011). *Practical Multivariate Analysis, Fifth Edition*, CRC Press. **(PMA5)**

Johnson, R.A. & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis, Sixth Edition*. Upper Saddle River, NJ: Prentice Hall. **(JW6)**

Other Books:

Abdelmonem A. Afifi, Virginia Clark & Susanne May (2003). *Computer-Aided Multivariate Analysis, Fourth Edition*, CRC Press. **(CAMA4)**

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning (Vol. 112)*. New York: Springer. **(ISL)** ; <https://www-bcf.usc.edu/~garth/ISL/>

Websites:

UCLA IDRE Textbook Examples for PMA 5th Textbook. <https://stats.idre.ucla.edu/other/examples/pma5/>
PSU Statistics Online course "Applied Multivariate Statistical Analysis":

<https://newonlinecourses.science.psu.edu/stat505/>

ETH Zurich Statistics Lectures "Applied Multivariate Statistics":

<https://stat.ethz.ch/lectures/ss18/applied-mvs.php/>

Course Requirements:

Homework:

There will be four homework assignments for this course, and each will be due at the beginning of the due date class. **Remember to include your code in the homework you turned in.** You may discuss with your classmates or work in teams for the homework, but each student has to submit their own homework. **Late homework will NOT be accepted without a reasonable and advance notice.**

Exam:

There will one take-home midterm exam. You are given one week to finish the exam. The format will be similar to the homework, however, **students CANNOT discuss midterm with their classmates.**

Project:

The project in this course accounts for a significant portion (40%) of the grade and should represent student's understanding of multivariate data analysis. Each student will be responsible for finding a real data set with numerous observations and variables, clearly stating the purpose of the analysis, carrying out the appropriate multivariate analysis relative to that purpose, **writing a report**, and **giving a 15-minute in-class presentation**. Grading will be based on both the report and the presentation.

There is no specific format requirement or page limit for the report. Generally, a formal report should at least include a title, the background introduction, the method applied and the major analysis results as well as the codes.

Course Website:

Course announcements, lecture notes, data sets, programs, and homework assignments will be distributed on the course webpage (<https://umd.instructure.com/courses/1261668>). Please check it on a regular basis. Lecture notes will be posted before class. You may wish to print these notes prior to each lecture and use them as an outline for taking notes during the class.

University Course Related Policies:

All University of Maryland-approved course policies are provided at the following website:

<http://www.ugst.umd.edu/courserelatedpolicies.html>

Policy descriptions, resources, and links to official policy documents are provided for:

Academic Integrity: What is cheating? What is plagiarism? What is the Honor Pledge?

Code of Student Conduct: What behavior is prohibited?

Sexual Misconduct: What to do in case of sexual harassment or sexual assault.

Non-Discrimination: Procedures to prohibit discrimination, complaints about discrimination, harassment, and retaliation.

Accessibility: Information about disability support services (DSS) and accommodations.

Attendance, Absences, or Missed Assignments: The student must notify the instructor in a timely manner (typically first week of class). Read this prior to Schedule Adjustment date.

Student Rights Regarding Undergraduate Courses: What should I find in the course syllabus?
Am I allowed to see my exams after they are graded?

Official UMD Communication: Use of email, communication with faculty, communication about cancelled class meetings, and weather-related or other urgent notifications.

Mid-Term Grades: Provided for 100 and 200 level courses, and all student athletes.

Complaints About Course Final Grades: Questions about course grades should first be addressed to the course instructor.

Copyright and Intellectual Property: Who owns the work that I produce in class?

Final Exams: Final exams are scheduled by the University.

Course Evaluations: The School of Public Health is committed to the use of student course evaluations for improving the student experience, course and curriculum delivery, and faculty instruction.

Campus Resources: ELMS, counseling, learning workshops, tutoring, writing help, questions about graduation, adding or dropping classes, withdrawing from the semester, etc.

Critical University Policies:

Inclement Weather / University Closings / Emergency Procedures:

In the event that the University has a delayed opening or is closed for an emergency or extended period of time, the instructor will communicate to students regarding schedule adjustments, including rescheduling of examinations and assignments due to inclement weather and campus emergencies.

Religious Observances:

The University System of Maryland policy provides that students should not be penalized because of observances of their religious beliefs; students shall be given an opportunity, whenever feasible, to make up within a reasonable time any academic assignment that is missed due to individual participation in religious observances. **It is the student's responsibility to inform the instructor in advance of any intended absences for religious observance.**

Special Accommodations / Disability Support Services:

If you have a documented disability and wish to discuss academic accommodations for test taking or other needs, you will need documentation from Disability Support Service (301-314-7682). If you are ill or encountering personal difficulties, please let the instructor know as soon as possible. You can also contact Learning Assistance Services (301-314-7693) and/or the Counseling Center (301-314-7651) for assistance.

Academic Integrity:

The University's code of academic integrity is designed to ensure that the principle of academic honesty is upheld. Any of the following acts, when committed by a student, constitutes academic dishonesty:

- CHEATING: intentionally using or attempting to use unauthorized materials, information, or study aids in an academic exercise.
- FABRICATION: intentional and unauthorized falsification or invention of any information or citation in an academic exercise.
- FACILITATING ACADEMIC DISHONESTY: intentionally or knowingly helping or attempting to help another to violate any provision of this code.
- PLAGIARISM: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise.
-

For more information see: <http://www.shc.umd.edu/code.html>.

The Honor Pledge is a statement undergraduate and graduate students should be asked to write by hand and sign on examinations, papers, or other academic assignments. The Pledge reads:

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination.

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit <http://www.shc.umd.edu/>.

Course Policies:

Absence Policy:

In accordance with University policy if you are absent for a single (1) lecture due to illness or some form of personal or family emergency, this absence will be considered “excused” and the instructor will accept a note from you attesting to the date of the illness/incident, along with an acknowledgement that the information is true. Whenever feasible, you should try to contact the instructor in advance.

Multiple or prolonged absences, and absences that prevent attendance at a major scheduled grading event (like an exam or test) will require written documentation from an appropriate health care provider/organization.

A link to pull information on the new policy covering absences from class can be found at <http://www.president.umd.edu/policies/v100g.html>.

Grading Procedures:

Grade of this course will be based on determined as follows:

- Homework 40%
- Midterm Exam 20%
- Project 40%

Course Outline / Course Calendar:

Tentative Course Schedule*			
Session	Date	Topic	Assignments
1	1/28/2019	Introduction to multivariate data: linear algebra basics, multivariate distribution, Hotelling's T^2 .	
2	2/4/2019	Multivariate analysis of variance (MANOVA) and multivariate linear regression	
3	2/11/2019	Principal components analysis (PCA) I: concept, interpretation, computation and visualization	HW1 assigned
4	2/18/2019	Principal components analysis (PCA) II: application, case studies and extension	
5	2/25/2019	Factor analysis I: model, estimation and interpretation; exploratory factor analysis	HW 1 due HW2 assigned
6	3/4/2019	Factor analysis II: application and case studies; confirmatory factor analysis	
7	3/11/2019	Classification I: concept, Bayes' rule, Discriminant analysis, classification evaluation	HW2 due
-	3/18/2019	Spring Break	
-	3/25/2019	No Class [#]	
8	4/1/2019	Classification II: other common classification methods and case studies	HW3 assigned
9	4/8/2019	Clustering analysis I: concept, distance measure, hierarchical clustering and K-means clustering	
10	4/15/2019	Clustering analysis II: model-based clustering, clustering	HW3 due,

		evaluation and case studies	Midterm assigned[^]
11	4/22/2019	Canonical correlation analysis (CCA) and its application	Midterm due, HW4 assigned
12	4/29/2019	Correspondence analysis (CA) for categorical data	
13	5/6/2019	Advanced topics: TBD	HW4 due
14	5/13/2019	Class Presentation	Project report due

* This is a tentative schedule, and the actual materials covered in each lecture might not be exactly the same.

Instructor will travel for a conference.

[^] Assigned date might change

Note: Numbers in brackets after learning objectives show linkage between material covered in each session and the numbered program competencies shown on page 1 of this syllabus.

Required Session Outline	
Session 1	1/28/2019
Topic: Introduction to multivariate data: linear algebra basics, multivariate distribution, Hotelling's T^2	
Learning Objectives for Session 1 [Relevant Program Competencies: #1, #3, #5]	
<ul style="list-style-type: none"> - Distinguish among the different measurement scales or types of variables and select appropriate descriptive statistical methods for summarizing public health data. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data. - Basic concepts of distance, basic linear algebra and multivariate distributions 	
Reading: PMA5 Chapter 7,8,9, JW6 Chapter 1,2,4	
Session 2	2/4/2019
Topic: Multivariate analysis of variance (MANOVA) and Multivariate linear regression	
Learning Objectives for Session 3 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]	
<ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. 	
Reading: JW6 Chapter 6,7	
Session 3	2/11/2019
Topic: Principal components analysis (PCA) I: concept, interpretation, computation and visualization	

<p>Learning Objectives for Session 3 [Relevant Program Competencies: #2, #3, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: PMA5 Chapter 14, JW6 Chapter 8</p> <p>Homework 1 Assigned (Due 2/25/2019 before class)</p>	
Session 4	2/18/2019
<p>Topic: Principal components analysis (PCA) II: application, case studies and extension</p> <p>Learning Objectives for Session 4 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: PMA5 Chapter 14, JW6 Chapter 8</p>	
Session 5	2/25/2019
<p>Topic: Factor analysis I: model, estimation and interpretation; exploratory factor analysis</p> <p>Learning Objectives for Session 5 [Relevant Program Competencies: #2, #3, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: PMA5 Chapter 15, JW6 Chapter 9</p> <p>Homework 2 Assigned (Due 3/11/2019 before class)</p>	
Session 6	3/4/2019
<p>Topic: Factor analysis II: application and case studies; confirmatory factor analysis</p>	

<p>Learning Objectives for Session 6 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: PMA5 Chapter 15, JW6 Chapter 9</p>	
Session 7	3/11/2019
<p>Topic: Classification I: concept, Bayes' rule, Discriminant analysis, classification evaluation</p> <p>Learning Objectives for Session 7 [Relevant Program Competencies: #2, #3, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: PMA5 Chapter 11, JW6 Chapter 11</p>	
Session 8	4/1/2019
<p>Topic: Classification II: other common classification methods and case studies</p> <p>Learning Objectives for Session 8 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: PMA5 Chapter 11, JW6 Chapter 11</p> <p>Homework 3 Assigned (Due 4/15/2019 before class)</p>	
Session 9	4/8/2019
<p>Topic: Clustering analysis I: concept, distance measure, hierarchical clustering and K-means clustering</p>	

Learning Objectives for Session 9 [Relevant Program Competencies: #2, #3, #5, #7, #8]	
<ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. 	
Reading: PMA5 Chapter 16, JW6 Chapter 12	
Session 10	4/15/2019
Topic: Clustering analysis II: model-based clustering, clustering evaluation and case studies	
Learning Objectives for Session 10 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]	
<ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. 	
Reading: PMA5 Chapter 16, JW6 Chapter 12	
Session 11	4/22/2019
Topic: Canonical correlation analysis (CCA) and its application	
Learning Objectives for Session 12 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]	
<ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. 	
Reading: Read PMA5 Chapter 10/ JW6 Chapter 10	
Homework 4 Assigned (Due 5/6/2019 before class)	
Session 12	4/29/2019
Topic: Correspondence analysis (CA) for categorical data	

<p>Learning Objectives for Session 13 [Relevant Program Competencies: #2, #3, #4, #5, #7, #8]</p> <ul style="list-style-type: none"> - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Reading: ED2 Chapter 4, JW6 Chapter 12</p>	
Session 13	5/6/2019
<p>Topic: Advanced topics: TBD</p> <p>Recommended reading: ISL Chapter 4,7,9,10</p>	
Session 14	5/13/2019
<p>Topic: Student in-class presentation</p> <p>Learning Objectives for Session [Relevant Program Competencies: #1, #2, #3, #4, #5, #6, #7, #8]</p> <ul style="list-style-type: none"> - Distinguish among the different measurement scales or types of variables and select appropriate descriptive statistical methods for summarizing public health data. - Select appropriate inferential statistical methods to answer research questions relevant to public health research. - Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research. - Critically evaluate statistical analyses presented in public health literature. - Use statistical analytical software packages (e.g. SAS, R, STATA) to describe, explore, and summarize data as well as perform statistical procedures. - Communicate results of statistical analyses to lay and professional audiences. - Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate. - Interpret results of data analysis for public health research, policy or practice. <p>Project report Due</p>	