EDMS 771 – Multivariate Data Analysis (Section 0101)
Department of Measurement, Statistics, and Evaluation (EDMS)
University of Maryland, Spring 2010 (3 Credit Hours)
Thursdays 4:15 – 7:00 p.m. Art-Sociology Building, Room 3207

Instructor: André A. Rupp
Office: EDMS 1230 A
Office Hours: Thur 2-3 & by appointment
Web: www.education.umd.edu/EDMS/fac/Rupp.html
Phone: (301) 405-3623
E-mail: ruppandr@umd.edu
Fax: (301) 314-9245

Course Overview

This course is a course on statistical models that are appropriate for the analysis of multivariate data. We will focus on the statistical theory underlying the statistical models as well on their application to real and, perhaps, simulated data sets using SPSS. Throughout the course, we will look at illustrative examples from a variety of disciplines that are included in the textbook and provided by me.

In the work that you will do for this course, you will frequently be asked to explain your reasoning or to think through a problem that does not have a single fixed solution. The main emphasis of this course is, thus, not merely to provide you with a toolbox for solving challenging psychometric questions, but to encourage you to become a reflective, principled, and thoughtful user of these tools. It is my firm belief that this will enable you to collaborate more effectively with colleagues, especially when working in interdisciplinary teams.

The joint prerequisites for this course are EDMS 645, 646, and 651 or equivalent courses. If you do not have these prerequisites, a permission of the instructor is required to take this course.

Textbook

Lattin, J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data.* Belmont, CA: Brooks / Cole.

E-mail

Should there be any unexpected changes to the course requirements or the course schedule due to unforeseen circumstances these will be communicated clearly in class and via email. In general, I primarily use the anonymous listserv for the course that is available to me to communicate with you. Since it automatically adds and deletes students who register for or drop the course, please inform me if you do not receive my emails. That may be a registration issue or an incorrect email address in the university database.

SPSS

In general, you may use any statistical package of your choosing to complete homework assignments or the take-home exam assignments. However, faculty in EDMS, the textbook, and I use SPSS versions 11.0 and higher for EDMS 645 and EDMS 646. Examples in class will come from SPSS for Windows version 14 or 16. There are essentially two possible courses of action for you for using SPSS.

Option 1: Buy nothing. Use a campus lab to do SPSS assignments for this class.

SPSS is available in the *Benjamin Building's computer lab (Room 0230)* in the basement of as well as in several computer labs on campus that you can find via http://www.oit.umd.edu/wheretogo/. You may use whichever SPSS version you wish, but you should know that (typically slight) differences may exist between what is presented in class and the options available on your chosen version.

Option 2: Buy either the student version or the GradPack version of SPSS.

Information about both versions can be accessed via the website http://www.spss.com/vertical_markets/education/SPSS_student_versions.htm The *GradPack* is the more complete version that does not have variable or case limits. It comes for up to 85% off the full price and there are currently two versions of it, one with just the base module for SPSS and one with add-on modules; only the latter would allow you to run almost all analyses that we are doing in this class. Buying the GradPack is a good investment if you plan to take more quantitative classes beyond this one and/or if you plan to conduct serious quantitative research (e.g., a Master's thesis or a dissertation). Note that SPSS is unfortunately not available as a student site-license via the university.

Attendance

Coming to the course not only ensures that you will be able to be exposed to the material in an alternative manner, but also ensures that you have an opportunity to engage in critical discussions with your classmates. Hence, attendance in this course is technically mandatory but since is a graduate-level seminar it is really your choice whether to attend this course; I do not record absences. However, should you be absent on a date when an assignment is due you are required to make arrangements for handing in the assignment by that day, which can most commonly be done simply via email or by putting the assignment in my EDMS mailbox.

E-mail

Should there be any unexpected changes to the course requirements or the course schedule due to unforeseen circumstances these will be communicated clearly in class and via email. In general, I primarily use the anonymous listserv for the course that is available to me to communicate with you. Since it automatically adds and deletes students who register or drop the course, please inform me if you do not receive my emails. That may be a registration issue or an incorrect email address in the university database.

Academic Accommodations

In compliance with and in the spirit of the Americans with Disabilities Act (ADA), I want to work with you if you have a documented disability that is relevant to successfully completing your work in this course. If you need academic accommodation by virtue of a documented disability, please contact me as soon as possible to discuss your needs. Students with documented needs for such accommodations must meet the same achievement standards required of all other students, although the exact way in which achievement is demonstrated may be altered. All requests for academic accommodations should be made as early as possible in the semester. For further information concerning disability accommodations, please contact Dr. William Scales at the Disability Support Service at phone number (301) 314-7682.

Academic Dishonesty
(Cheating, Fabrication, Plagiarism, Facilitation)

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 to uphold these standards for this course. It is imperative that you are 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.studenthonorcouncil.umd.edu/code.html for details. Plagiarism and other forms of academic fraud are a violation of university regulations and unacceptable under any circumstance. This means for this course specifically that instances of duplicating either answers on the homework or sections of the final paper have to be and will be reported to the Honor Council in writing.

Course Evaluations

Your participation in the evaluation of courses through *CourseEvalUM* is a responsibility you hold as a student member of our academic community. Your feedback is confidential and important to the improvement of teaching and learning at the University as well as to the tenure and promotion process. *CourseEvalUM* will be open for you to complete your evaluations for fall semester courses between Tuesday, April 27 and Wednesday, May 12. You can go directly to the website (www.courseevalum.umd.edu) to complete your evaluations. By completing all of your evaluations each semester, you will have the privilege of accessing the summary reports for thousands of courses online at *Testudo.*

Assessment Policy

Assessment will be in accordance with university policy, which states that "An instructor is not under obligation to offer a substitute assignment or to give a student a make-up assessment unless the failure to perform was due to an excused absence, that is, due to illness (of the student or a dependent), religious observance (where the nature of the observance prevents the student from being present during the class period), participation in university activities at the request of university authorities, or compelling circumstances beyond the student's control. Students claiming excused absence must apply in writing and furnish documentary support for their assertion that absence resulted from one of these causes."

Any changes to the assessment procedures or guidelines will be communicated clearly in class and via email.

Assessment Components

1. Homework Assignments (4 × 10 points)	40 %
2. Statistical Critique Papers (2 × 7.5 points)	15 %
3. Take-home Exam (1 × 25 points)	20 %
4. Course Paper (1 × 25 points)	25 %

Grading Scale

A+:	97-100	**B+**:	85-87	**C+:**	75-77	**D+:**	58-65	**F:**	< 50
A:	93-96	**B**:	81-84	**C:**	70-74	**D:**	55-57		
A-:	88-92	**B-:**	78-80	**C-:**	66-69	**D-:**	50-54		

Quality of Academic Writing and Presentation

One of the fundamental skills required to be successful in this class is critical and reflective thinking, which is in part developed through writing, because writing is the process of knowledge creation and transformation. Hence, I expect that your written work – especially the statistical critique papers and the course paper – satisfies the following criteria:

(a) Arguments are logically consistent and internally coherent
(b) The choice of words, phrases, and expressions is appropriate for formal academic American English as used in university and research contexts
(c) The grammar is correct and there are no typographical errors
(d) The work is word-processed and printed on clean paper with numbered pages that are stapled together
(e) The organization of the final paper follows the common conventions for a paper of this type unless deviations can be clearly justified due to the novelty of the methodological approach.

I reserve the right in this course to deduct points for all assignments if your written work is not prepared professionally and / or if the language or organization is so poor that your arguments are difficult to comprehend.

Assessment Descriptions

Homework *(4 × 10 points = 40 points)*

I believe that it is important for all students to continually attempt to integrate novel information with previous learned material for a deeper learning, which first of all requires constant practice and proper feedback. Hence, each class I will give you several homework assignments that you will prepare for the assigned due dates. I will typically return your homework the following week when we will also discuss it in class. The homework tasks will include doing a few computations by hand, using software to analyze data, as well as critiquing a provocative question, statement, or reading. It is important to me that arguments and thoughts are always critically evaluated and presented in a clear, concise, and coherent manner. In alignment with the preceding, I expect you all to hand in your homework in a professional manner, which implies that sentences are well-formed and logically consistent, work is presented cleanly with answers numbered successively and clearly, and is handed in on clean paper without things crossed out, smeared over, or otherwise messy. Late homework without any prearrangement is not acceptable. However, I am generally flexible if you need an extra day due to events outside your realm of control such as illness in alignment with the attendance policy for this course - if you inform me of your reasons beforehand. I will not accept last-minute requests for an extension via email or phone or explanations after the due date for work that was handed in late.

You are encouraged to work with your classmates on your homework assignments and can hand them in with at most one additional person in the class. Should you choose to do so, both of you will receive the same score.

Statistical Critique Papers *(2 × 7.5 points = 15 points)*

There are a total of two statistical critique papers in this class, each one based on one source that is selected by me. The sources will be journal articles, conference presentations, technical reports, or white papers that are available online. The purpose of the statistical critique papers is for you to comment on the manner in which different modern measurement models are applied in practice and the appropriateness with which such applications are documented. As a rough guideline, I would expect that each critical review paper would be about 3-5 pages long when you use 1.5 line spacing and 12-point Times New Roman font. Each statistical critique paper should consist of two sections. In the first section, you will summarize (a) the general context of the study, (b) the research question(s) that are addressed by the analysis, (c) the key variables that are used in the analyses, (d) the models that were used to analyze the data, and (e) the interpretations that were drawn from the analysis in order to answer the research questions. In the second section of each critique paper, you will critically evaluate the appropriateness of the interpretations by considering aspects of the studies including, but not limited to, whether model assumptions may have been violated, whether model fit has been assessed, whether bias analyses were conducted, whether effect sizes are reported, and whether the analysis was most appropriate for answering the targeted research question. You will also comment on whether the methodology and the results were presented in a clear manner for the intended audience. Finally, I would like you to provide recommendations for how the descriptions in the source could be improved pretending you are an editor and the material was submitted for publication.

You are encouraged to work with your classmates on the critique papers and can hand them in with at most one additional person in the class. Should you choose to do so, both of you will receive the same score.

Course Paper *(25 points)*

For the course paper, I will give you an option depending on your skills and interests as well as the availability of suitable data and commercial software packages to you. You can either choose to write an empirical analysis paper, which has, at its core, the analysis of a real-data set with appropriate multivariate modeling techniques. Alternatively, you can choose to conduct a small-scale simulation study that numerically investigates statistical properties of selected multivariate modeling techniques.

You are encouraged to work with your classmates on the course paper and can hand it in with at most one additional person in the class. Should you choose to do so, both of you will receive the same score.

Option 1: Empirical Analysis Paper.

This paper is a useful paper in order to fully appreciate the complexity of a modeling process. If you choose this type of paper, I would like you to develop methodologically and / or substantively relevant research questions, research what has been recently done to answer such questions in the domain of interest, find a data set appropriate for answering the research questions, and determine a set of models appropriate for the purpose of the study. You will then fit this suitable set of models to the data, report the relevant outcome statistics such as indices of global and relative model fit, item and / or person fit, and model parameters of a best-fitting model, and interpret the results in the context of the problem.

Specifically, the paper should illustrate…

(a) why the research questions you have chosen are relevant to the field of study
(b) how you justify your methodological approach to answering these research questions
(c) how you have empirically checked for the assumptions of the models you use
(d) how you have empirically fitted alternative models and selected certain final ones
(e) how you interpret the model parameters with respect to the research question of interest, and
(f) how you would characterize the limitations of your analysis and suggested directions for future research.

The paper needs to contain a few peer-reviewed references (i.e., articles and book chapters from peer-reviewed journals and books, research reports and technical papers, papers presented at peer-reviewed conferences) relevant to the domain that the data set is embedded in as well as a few references relevant to the modeling approaches you select to provide some contextualization for the reader.

The paper needs to be in proper APA format with appropriate title page, abstract, headings, citation format, and reference list. Please consult the most recent APA guide for details. As a rough guide, I would expect that your paper will be about 10-15 pages in length when you use 1.5 line spacing and a 12-point Times New Roman font.

Since this paper requires you to use a real data set you will need to determine an appropriate data set early on in the term yourself. Do not wait until the end of the term, because it is your responsibility to find it!

Option 2: Small-scale Simulation Study.

If you are more inclined to investigate the statistical properties of selected multivariate modeling techniques, you are invited to conduct a small-scale simulation study that investigates some of their statistical properties. In order to fully articulate your motivation of the simulation study, you will need to develop methodologically relevant research questions, research what has been recently done in this area, determine a suitable design for the simulation study, implement the design by either writing your own code or using routines in existing software programs, summarize the results using suitable outcome measures, and interpret the findings with respect to the research questions at the outset.

Specifically, the paper should illustrate…

(a) why the research questions you have chosen are relevant to the methodology
(b) how you justify your methodological approach to answering these research questions
(c) how you have selected the factors of the simulation study and their associated levels
(d) why you have chosen to combine the factors in the particular way that is your design
(e) how you have selected and computed the outcome measures
(f) how you have chosen to summarize the outcome measures using secondary methods
(g) how the results from the study address the research questions of interest, and
(h) how you would characterize the limitations of your analysis and suggested directions for future research.

The paper needs to contain a few peer-reviewed references (i.e., articles and book chapters from peer-reviewed journals and books, research reports and technical papers, papers presented at peer-reviewed conferences) relevant to the methodology that you are investigating to contextualize the work for the reader.

The paper needs to be in proper APA format with appropriate title page, abstract, headings, citation format, and reference list. Please consult the most recent APA guide for details. As a rough guide, I would expect that your paper will be about 10-15 pages in length 1.5 line spacing on 12-point Times New Roman font and that it would contain a more extended appendix than the empirical analysis paper.

Since this paper requires you to set up and conduct a small-scale simulation study, make sure to choose a design whose implementation is feasible within the time constraints of the course!

Take-home Exam *(20 Points)*

The take-home exam covers material that was taught up to one week before the due date and consists of a series of constructed response, computation, and interpretation questions. You will receive the exam on Thursday, March 11, and will have until Thursday, April 2, to complete it.

Contrary to the other assignments in this course, you are NOT allowed to work with anyone on the take-home exam and have to hand it in individually!

Based on the University policy and the short time frame in the summer, make-up exams will not be given. For your information, the University policy states that "an instructor is not under obligation to offer a substitute assignment or to give a student a make-up assessment unless the failure to perform was due to an excused absence, that is, due to illness (of the student or a dependent), religious observance (where the nature of the observance prevents the student from being present during the class period), participation in university activities at the request of university authorities, or compelling circumstances beyond the student's control. Students claiming excused absence must apply in writing and furnish documentary support for their assertion that absence resulted from one of these causes."

Additional References

The following is a list of additional sources that you may consult for the topics we discuss in class; note that there are also a large number of multivariate statistics textbooks on the market – in addition to Tabachnick and Fidell (2006) – that you might find useful also.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: Wiley.

Borg, I., & Groenen, P. J. F. (2009). *Modern multidimensional scaling: Theory and applications*. New York: Springer.

Gan, G., Ma, C., & Wu, C. (2007). *Data clustering: Theory, algorithms, and applications*. Alexandria, VA: American Statistical Association.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis.* New York: Cambridge University Press.

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis.* New York: Wiley.

Jackson, J. E. (2003). *A user's guide to principal components.* New York: Wiley.

Kline, R. B. (2004). *Principles and practice of structural equation modelling* (2nd ed.). New York: Guilford Press.

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (2004). *Applied linear statistical models* (5th ed.). Chicago: Irwin.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). New York: Allyn & Bacon.

Schedule and Reading List

Date	Topics	Textbook Chapter	Assignment Due
Jan 28	Introduction to Multivariate Data Analysis Motivations for multivariate analyses; multivariate general linear model	1	---
Feb 4	Matrix Notation for Multivariate Data Analysis Basic components; regression models in matrix notation	2	---
Feb 11	Logistic Regression Models Simple and multiple logistic regression; dichotomous vs. polytomous models	Supp 1	HW 1
Feb 18	Log-linear Models Main effect and interaction models; relation to logistic regression models	Supp 2	---
Feb 25	Multivariate ANOVA Main effect and interaction models; relation to experimental design	11	HW 2
Mar 4	Principal Components Analysis Basic model; selection of components; relation to other data-reductions	4	---
Mar 11	Exploratory Factor Analysis Basic model; uni- vs. multi-dimensional models; relation to PC analysis	5	HW 3
Mar 18	**SPRING BREAK** **NO CLASS**	---	---
Mar 25	Confirmatory Factor Analysis Basic model; model selection; relation to exploratory factor analysis	6	---
Apr 2	Multidimensional Scaling Basic model; relation to other multidimensional scaling approaches	7	Take-home
Apr 8	Cluster Analysis Distance metrics; K-means approaches; hierarchical approaches	8	Critique 1
Apr 15	**NCME & AERA** **NO CLASS**	---	---
Apr 22	Latent Class Analysis Exploratory vs. confirmatory models; relation to other grouping methods	Supp 3	HW 4
Apr 29	Canonical Correlation Analysis Basic model; model selection; relation to other regression-type methods	9	Critique 2
May 6	Discriminant Function Analysis Basic model; model selection; relation to other classification approaches	12	---
May 13	**OFFICIAL FINAL EXAM DAY** **NO CLASS – COURSE PAPER DUE**	---	Course Paper

Note. The supplementary readings will be made available electronically via email and on the ELMS blackboard.