Course: 07.706 Intermediate/Advanced Data Analysis
Time: Tuesday 4:00 to 6:30        Room: As assigned

Instructor: Dr. James Carifio       Office: 527 O’Leary Library
Phone: 978-934-4609   Email: James_Carifio@uml.edu
Office Hours: Mon, Tues, and Thurs, 6:30-7:30, and by appointment and e-mail.

General Course Overview
The field of intermediate and advanced data analysis is more than very extensive and complex, and particularly so when one includes advanced qualitative techniques (including content and non-parametric analyses) and mix methods approaches (which are actually multivariate by definition). Along with the classic and traditional quantitative methods and the areas of change assessment, psychometrics, research synthesis, and theory development and testing, multivariate analyses are also used in a broad variety of ways ranging from the analysis and comprehension of text by computers, to decision-making in various kinds of “smart” software, to data mining the web or “corporate” database, to analyzing genomes and characteristics of extant species of organisms, and to analyzing other large, complex, and amorphous data sources such galaxies. Multivariate statistical techniques are very powerful and generalized tools that may be used to analyze a broad array of data and research designs in ways that avoid the numerous pitfalls of univariate analyses of the same data and designs and to uncovering effects and findings that are missed by univariate techniques. In general, multivariate analyses are about analyzing collections of variables (complexes and complexity) reliably and validly in terms of (real) differences, effects, basic structures and causal relationships. Multivariate analyses are sometimes called “data reduction techniques” in terms of seeing and understanding the “simple” overall picture in the complexity (of variables) being considered. There is no way that all of these areas can be cover in one course, as many of these methodologies, approaches and analytical techniques would take an entire seminar on their on to learn and understand at a professional level as well as to become adept at using them in nuanced and sophisticated ways. Therefore, like all multivariate analyses we must first seek to simplify this complexity.

Students come to this course which many different foci and many different needs, and the course will be built around these foci and needs so that students may become most proficient in the analytical procedures and skill sets that they most need for their own individual work. However, to understand and be able to interpret and use any one technique or model of analysis (e.g., content, trend, factor, logit, or path analysis), there is a ‘core curriculum’ and ‘a general core set of knowledge, understanding and skills’ that one must have and have mastered.

The first third to half of this course, therefore, will be spent on this core curriculum. This strategy is used to make sure that everyone understands this core and its key concepts, ideas, facts, procedures, critical technical points, and basic theory as this core underlies all of the techniques available to be studied. This core also includes a detailed and in-depth understanding of variances, correlation, measurement errors, sampling errors, types of relationships and types of change and how they related to complex rather than (overly) simple data sets. Multivariate data and analysis is just that: appropriately conceptualizing, analyzing and interpreting complexity. For some, the first part of the course will be review to some degree. For others, it will be new learning. As we do this ‘core curriculum,’ we will also do a ‘general tour’ of the intermediate and advanced multivariate methods that are in existence at the conceptual level and what they do and achieve as an analytical strategy.

Often to do a study or to analyzed a complex dataset, several different multivariate techniques have to be used and used sequentially (e.g., factor analyzing the data before doing a trend or multiple regression analysis), so there is a “putting it together” phase (and tour) too. Once we have done the basics, we will choose some topics,
techniques, or procedures to focus on in greater depth in the second half of the course. Depending on the group, this could lead to small group work, independent work, and class time allocated to specific topics, groups or individuals (well known in advance to everyone) as opposed to everyone necessarily attending all of the class all of the time, although everyone may do so as much can be learned from other people’s topics, problem and struggles. If you chose not to attend a class or part of a class, you should be using the missed class time wisely to carry out you individual assignments or doing library research work on it. This factor is one of the reasons why I do not require class attendance in this course (see below for details).

I have assembled and will make available to you in one form or another a wide variety of resources which include on-line sites, books and articles, and other relevant texts, articles, and handouts as well as various bibliographies. Note well: www.Albris.com is a source of quickly delivered and relatively cheap texts and books so asking you to have several core books if you need them is not an onerous burden if you cannot obtain them from other sources.

For the core component of the course, the Mertler and Vannata text (Pyrczak Publishing, 1-884585-59-0) at the Book Store and a good univariate statistics book (see list and ogo to www.Albris.com if you do not have one) along with a basic research methods book (see list), should suffice, particular with all of the on-line statistics books and sites I will give you.

For the second half of the course, I will tell you what books to get to focus on a group special topic or topics, if there is one such as meta-analysis. However, if the topic(s) chosen is not your interest or focus for the second half of the course (other than to learn how to read a meta-analysis critically and with understanding), you will have to acquire instructional resources from those given on various lists to have appropriate support materials for yourself. If special topics and individual foci do not emerge and develop, we will follow the standard default course topics schedule given at the end of this syllabus. I have explained this point to you here so you will understand other parts of this syllabus.

Assessment and Grading

There will be four take-home examinations/projects in this course equally spaced. The first two will be on the basics of the “core curriculum” (20% each). These two take-home exams will be a mixture of short answer/essay questions and interpretation of analyses and results type items. The third take-home exam will on the basics of your chosen analysis topic area (30%) and the last (30%) will be either your summary of the literature on your chosen analysis topic or your project analyzing some data using your chosen analytical procedure and writing your analyses and results up. Extra credit of 15% will be given for doing weekly in and out of class assignments and this extra credit will be used to smooth out variability in your 4 take-home exam assignments.

A criterion-referenced grading system will be used so it is possible for everyone to do well or poorly. FYI, most students do well in this course even though is hard, demanding, complex, and exacting but exceedingly interesting because it becomes obvious to students that they are acquiring powerful analytical and evidence processing skills and understandings, and can read the current educational literature more critically and with greater understanding. If topic areas do not emerge and the “default curriculum” is pursued, then each exam will follow the default curriculum and have a value of 25%. The ‘extra credit’ 15% factor will also be used for the default curriculum.
Basic Resources for Course:

**FREE-ON-LINE Stat books, software and helpful web sites:** You will need to email me at the following email address ([James_Carifio706@yahoo.com](mailto:James_Carifio706@yahoo.com)), and I then email you back an electronic form of the list of on-line books and sites I will hand out in class, so you will not have to type them all in. Emailing me will start the process of your receiving electronic handouts from me and post-class follow-up memos as well as my answering your questions and your turning in certain assignments to me. The Yahoo address is the address for doing “course business” to keep it all in one place and my university email uncluttered.

**The Core Text:** is Mertler, C and Vannatta, R. *Advanced and Multivariate Statistical Methods* (Pyrczak Publishing, 1-884585-59-0) at the South Campus Bookstore. Excellent basic, minimally mathematical text that covers core topics and methods including the SPSS commands for selecting analyses and explanation of the SPSS output from analyses. This text has a website where you can download the data used for the analyses in the book and other features. There are numerous other websites associated with multivariate analyses and texts (see lists).

**SPSS (Statistical Programs for the Social Sciences) software and instructional texts on how to use this software package:** A separate handout will be given on SPSS (and other statistic software packages/options) and the variety of texts available for learning to use it. The Student version of SPSS may be obtained by going to [www.SPSS.com](http://www.SPSS.com) as can helpful material on SPSS and how to use it for those students who want to jump into the deep end ASAP. The student version of SPSS is fairly comprehensive and does most of the basics in multivariate analysis. FYI, all versions of SPSS has a tutorial program and an on-line help program. FYI, the versions of SPSS available in the computer labs at UML only have a few of the multivariate functions (or did so the last time I checked). FYI, older versions of SPSS are usually Ok for the actual analysis modules the version has as the newer versions are additions and changes other the actual specific statistical analyses in the version other than adding new ones not present in the older versions such as content analysis or survival analysis and so on.

**INSTRUCTOR HANDOUT MATERIALS:** supplements to and expansions of the above content as well as many articles to read on specific issues and problems including examples of specific analyses and combinations of analyses and data from real research projects for you to analyze and compare you results to the author’s.

**Supplemental Resources**

**DOCTORAL LEVEL BASIC STAT BOOKS:**

The *Lowery On-line Text* is an excellent doctoral level text and one of the ‘core texts’ for this course, as is a copy of George Ferguson’s *Statistical Methods (3-rd to 5-th editions)*, which is out of print, but may be obtained at various libraries or at [www.Libris.com](http://www.Libris.com). There are other similar texts (see below).

Ferguson, George and Takane, Yoshio. *Statistical Analysis in Psychology and Education*, Sixth Edition (Alibris.com, Library, local libraries ,other students). Also excellent doctoral level stat book but has some errors not in the older additions. Is more complete on some of the last multivariate and item analysis chapters than the older editions.
Glass and Hopkins. **Statistical Methods in Education and Psychology.** In print, but an excellent detailed doctoral level basic stat book.

Shavelson, R. **Statistical Reasoning in the Behavioral Sciences.** In print, widely used doctoral level stat book but with some flaws.

There are many good basic stat books (see anything written by Allen Edwards). But remember no one stat book is excellent or even good on all things, topics, points, or issues. Every book is good on some things, not good on others, and has important things missing, as the book reflects the author’s and publishers views and choices. So typically, you need to be familiar with several good sources that you consult depending upon the issues, problem or point under consideration. So, please do not tell me that something I said or we covered in class was not in “THE BOOK,” which tends to mean the ONE book you have chosen to read as opposed to the books you should be familiar with as basic resources.

**RESEARCH DESIGN BOOKS**

Multivariate Analysis not only assumes you know the basics of inferential statistics (both parametric and non-parametric) but is also assumes that you know the basics of experiment, quasi-experimental and non-experimental design which you should as you have had the 702 course (big smile here). You can do a quick review of this material with the books you had in your 702 course as well as with Campbell and Stanley’s classic **Experimental and Quasi-Experimental Design** (can be read in a weekend) and the Campbell and Cook more detail version by the same title (both available at www.Albris.com). Depending on what multivariate analysis you do and what areas you might get into (e.g., fractured factorial ANOVAs), you may need a more sophisticated experimental design book such as David Keppel’s **Design and Analysis of Experiments** or Kirk’s **Design of Experiments** or something similar that is on the bibliography list. For some of the more sophisticated design’s I have handouts but you need to know the basic ones at a minimum and the now 20 threats to internal and external validity (which we will review). You will also need to know some psychometrics and reliability theory in particular and validity theory. Kerlinger and Lee’s **Foundations of Behavioral Research** is excellent in this regard and Chum Nunnally’s **Psychometric Methods** is one of the bibles in this area. Because I have found a quick source of cheap books (comparatively) for you, I expect you to get books you might need when you need them if you cannot obtain them elsewhere (i.e., the library, or friends).

**Ground Rules and Your Responsibilities**

I have a separate handout on the ground rules for this course and your responsibilities that we will go over together point-by-point as a class in the first meeting. In general, I expect you to be an independent, self-directed, self-managing, adult learner because that is what you are going to need to be as a doctoral student and most definitely to complete a dissertation. I will provide you with many different kinds and types of resources and support, but it is you who must make all of the requisite and prudent efforts to learn all of the things you need to learn to be successful in this course. I will also try to give you (meta-cognitive) tips and guidance throughout the course to help you.
For example, the most basic thing you need to learn first about statistics is that it is a highly cumulative and highly inter-related discipline (set of concept, principles, facts, theories, kind of knowledge etc), which means that you must work on learning it every day or every other day and if you do not learn and understand something, it is going to come back and bite you and bite you hard usually. You must be a “plodder” with this content and do it slowly, carefully and systematically, and particularly so with multivariate analyses.

You cannot cram this content, particularly if it is new and unfamiliar to you. You cannot cursory or half-learn it. You must learn it with precision, depth and detail. All of this takes time and being methodical. It is like getting into shape or dieting. There are no miracle pills or short cuts and some of it is basically boring as all get out (but become very exciting later) and takes true grit and stamina most of the time. So it is your responsibility to do the readings, exercises, problems and other assignments BEFORE YOU COME TO CLASS and do them in a systematic and methodical manner rather than in a last minute cramming way, or not at all (and I am famous for surprise quizzes or surprise in-class-problems). This particular responsibility is the most basic of your responsibilities in this course.

If you have not done the readings and exercises in a systematic and methodological fashion before coming to class, or have a one-on-one meeting with me, you will have broken covenant with me. And, I will ask you for “PROOF-OF-EFFORT” if I think you have not been doing the assigned work. If you cannot provide me with proof of effort, please to not expect me to be supportive or understanding of you, or to compensate for your lack of required effort relative to your meeting your basic responsibilities in this course. If, on the other hand, I see observable proof of effort, I will (within reason) keep trying to find ways to help you understand this content and develop the skills you need to develop in this course.

You also need to learn the STOP-RULE: if you’ve worked on a problem for 20 minutes and are lost and flummoxed STOP AND GET HELP; do not work at it for another hour. This same rule applies to concepts: if you do not have a clue after an hour trying to figure it (and reading alternative sources): STOP AND GET HELP (which includes contacting members of the class or people you know); do not work on it for another two hours and then vent at me in class about the chapter or the assignment.

Lastly, most of you will not fully (or even minimally) understand the (matrix) mathematics of multivariate analysis (and particularly in the short period of time that we have for this course). Most people do not and that is a plain and simple fact, so do not get upset, embarrassed, and ashamed or frustrated by this fact. Just accept it, as it will not stop you from being a good multivariate analyst if you know all of the assumptions of a given test and check that each one is met, know the major ways in which the test can malfunction, be biased or produced misleading or erroneous results and check if any of these malfunctions have occurred, have an understanding of what valid results should in general look like across the complexity of the analysis, and know how to read and interpret each piece of the results that are part of the analyses that are done by the software you use and how to do the ones that the software does not due if they need to be done to complete the full picture before interpreting the results and drawing conclusions. If you know and do these things, you will be an excellent multivariate analyst as the math has been gone over and checked “to the nines” and has been and is continually being research. In a word, you do not have to be a quantum physics or neurology expert to use a chisel and hammer and sculpt a David. You just have to “know your tools” extremely well and the nature of the “stone your are working with” and how to analyze and evaluate the results that are occurring as you execute your analytical plan that is the results of your thinking, imagination and design.
I can give you no more practical or wise advice than the above, particularly if and when you start to work independently on your own specialty area. But if you follow this advice you will find that by the end of the course that you will have acquire some very powerful analytical skills and evidence processing capabilities that will make you a very different kind of educational analyst and stand you in good stead in many different ways over the course of your professional career.

**Conceptual Framework: Education for Transformation**

The central tenets of our conceptual framework are

*Excellence, Equity, Inquiry and Collaboration*

One of the core skills one needs for evidence based practice is the ability to generate and analyze reliable and valid evidence and to critical review and evaluate the evidence and analyses generated by others and particularly quantitative evidence and analyses. One must be able to critically read and interpret a broad array of evidence and evidence based literature and studies and be knowledgeable about the flaws, shortcomings, pitfalls and errors of various statistics, methods, designs, and analyses. One must be able to critically evaluate the evidence based arguments of others and one must be able to engage in professional and scholarly evidence-based discussions, dialogues, debates, and problem solving activities with others as well as to properly write-up and present the results of one’s own evidence-based inquiries. Inquiry and research as well as both the science and art of evidence and its interpretation has established standards of excellence and high quality evidence is a key component to in the pursuit of educational and personal excellence and achievement. Learning to analyze evidence in terms of individual differences and from multiple perspectives and theories is critical to full and appropriate analyses and to the pursuit of equity, whether it be personal, intellectual, or and social.

Inquiry and research is by its very nature collaborative in its conception, design, review, conduct, sharing, intersubjective validation and translation into practice, and what is called the research literature is a collaborative community that is both nurturing and sustaining once one become an active and literate member. A variety of literature in several disciplines shows that discoveries are made and problems are solved more by teams today than by individuals. Various new studies of publication authorship show that the quality of research and scholarly articles written by multiple authors tends to be higher than research and scholarly articles by single authors and that roughly 70% of scholarly publications in upper tier journals today tend to be multiply authored. Research and scholarly articles in upper tier journals are reviewed by multiple reviewers who differing views of the article must be reconciled before the article is published. And, of course, your dissertation committee will have a minimum of 3 professors on it for a variety of reasons. Learning to dialogue professionally and collaborate and work with your colleagues and peers and the literature in your own and related areas are key sets of achievements and skills you need today to be a capable professional as well as to do high quality and excellent research and scholarship, as the current research shows.

In this course, you will develop the core knowledge and skills you will need to generate, analyze and interpret reliable and valid quantitative and qualitative evidence and to critically read and evaluative the same generated by others. You will learn the cannons and standards of excellent evidence as well as the limitations of all evidence. The use of a criterion reference assessment and grading system in this course is to give you first hand experience with a cannon and standards as are the responsibilities I have outlined relative to your own learning. You will learn how to generate and analyze evidence that allows individual differences to be considered and addressed and allows differing views to be assessed and tested; both are equity concerns. You will also learn the norms, cannon, and etiquette of scholarly and professional collaboration relative to evidence and research (by the time you finish the second course required in this area). My use of examples, simulations, and research vignettes for you to analyze and interpret as well as actual data and my use of objective assessments are all directed at these goals as well as my use of learning and support groups in this course.
The Default Schedule/Syllabus: M&V = Mertla and Vanatta text

Sept 9: Overview and review of basic statistical and research design concepts including the concept of effect sizes. Differences between univariate and multivariate analyses, and parametric and non-parametric analyses. Individual interests and need assessments. Tentative Plans. Measurement, reliability and validity (of measures and of study designs) review. Follow-up Reading: assigned in class and by email.

Sept 16: Elaboration and clarifications of points from previous week and assigned materials. Introduction to multivariate statistics and a guide to multivariate techniques. Reading: M&V: Ch 1 and 2.

Sept 23: Data Preparation and Pre-data Screening and checking (neither can be over-emphasized). Statistical software options and strengths/limitations. ANOVA review. Reading M&V: Ch 3 and 4 and handouts. 
Take-home Exam 1

Sept 30: Data reduction, underlying structures, instrument development and validation and reducing complexities: Factor Analysis the first time. Reading M&V: Ch 9 and handouts.

Oct 7: Elaboration and clarifications of points from previous week and assigned materials. Analysis of Covariance. M&V: Chapter 6 and handouts. Return Take-home Exam 1

Oct 14: Multivariate Analysis of Variances (MANCOVA) and Multivariate Analysis of Covariance (MANCOVA). Readings: M&V Ch 6 and handouts.

Oct 21: Multivariate Analysis of Variances (MANCOVA) and Multivariate Analysis of Covariance (MANCOVA). Readings: M&V Ch 6 and handouts. Take-home Exam 2

Oct 28: Multiple Regression the first time. Readings M&V: Ch 7.


Nov 11: Multiple Regression the second time. Reading: Handouts. Take-home Exam 3.

Nov 18: Discriminant Analysis (the under utilized technique and conceptual model). Reading: M&V Ch 10.


Dec 9: Cluster Analysis and Cannonical Regression. Readings: Handouts
