

---

From John C. Paolillo

# Teaching: S604s08

---

SLIS S604 Topics in Library and Information Science

## Statistics for Information Science and Usability

John C. Paolillo, Instructor ([paolillo@indiana.edu](mailto:paolillo@indiana.edu))

Thursdays, 5:45-8:30 PM, LI 002

Office	Hours
LI 030A	R 4:00-5:30
Eigenman 912	TBA

### Quick Links

- [PDF version of syllabus](#) <sup>Δ</sup> (current 1-15-2008)
- [Course wiki](#) discussion of readings, assignments, etc.

### Course Description

Information Science and Human-Computer Interaction depend heavily on empirical research in their ways of creating knowledge. Empirical methods depend in turn on statistics, which offers a number of different approaches and techniques for extracting patterns from observed phenomena. Hence, people who wish to conduct research in these fields can benefit greatly by obtaining exposure to these approaches and learning how to apply them to answer a variety of research questions.

This course offers an overview of multivariate multivariate statistics, and its applications to the research questions of Information Science and Human-Computer Interaction. It assumes no prior knowledge of statistics, developing the notions of distribution and model through example analyses drawn from the research literature of the fields of Information Science and Human-Computer Interaction, using original source data where possible.

Two types of statistical models are emphasized: Generalized Linear Models, and latent structure models. Generalized Linear Models include and extend the linear model family (ANOVA, linear regression) to situations where counts or proportions are analyzed. Such models are often needed in a range of information science and usability research. Latent structure models include Principal Components, Factor Analysis, Multi-Dimensional Scaling and other forms of vector spaces, as well as forms of cluster analysis. These

models are important for data exploration and visualization, as well as in information retrieval system design and performance evaluation.

The principal software package used in this course is R, a free (Gnu Public License) statistical programming language and environment, used in applications from investment banking to social network analysis to bioinformatics. R offers a highly flexible, command-line environment that is robust enough to handle very large data sets, with facilities for producing high-quality graphics.

## Prerequisites:

SLIS L509/709 or permission of the instructor is a prerequisite for this course. In general, prior college-level math skills, and exposure to social science research methods should be sufficient preparation.

## Course Objectives

This course aims to develop an understanding of multivariate statistical models and how they are applied to research in information science and usability studies. At the end of this course, students should be able to

- Identify a range of statistical models and describe their relationships to one another
- Select an appropriate statistical model for addressing research questions about a particular data set
- Conduct a range of analyses using R
- Interpret the results of the statistical analyses
- Critically interpret the results of statistical analyses presented in research literature presented in information science and human-computer interaction

## Requirements:

The course aims are achieved through:

- Readings from information science and human-computer interaction literature in which statistical data are presented
- Demonstrations and hands-on example analyses using data from information science and usability studies
- Assignments in which statistical analyses are run on data from actual research studies

The graded work in this course comprises five assignments worth ten points each, a final presentation, worth 15 points, and a final paper or project worth 45 points. The topic of the final presentation and the final paper or project are the same, and will be chosen by each student in consultation with the instructor.

## Calendar

The following calendar outlines our activities for the semester. This portion of the syllabus, in particular, may change, so please check it regularly. You are responsible for readings on the day indicated.

Date	Readings	Assignments
Jan 10	Introduction, syllabus	
Jan 17	MSA 1: Fundamentals (measurement, scales, etc.)	
Jan 24	MSS 2, 1st half: data	I
Jan 31	MSA 2: Probability	
Feb 7	MSS 2 (to end): Data	II
Feb 14	MSS 3 (to p.68), MSA 5: ANOVA	
Feb 21	MSS 3 (to end), MSA 4: Regression (Ordinary Least Squares)	III
Feb 28	MSS 4: Logistic Regression	IV
Mar 6	MSS 5: Log-linear modeling	V
Mar 13	Spring Break	
Mar 20	MSS 6, MSA 7: Factor Analysis	VI
Mar 27	MSA 8: Cluster Analysis	VII
Apr 3	MSA 9: Multidimensional Scaling	
Apr 10	Overview of multivariate statistics	VIII
Apr 17	(open day)	
Apr 24	Presentations	
May 1	Final Papers due	

## Textbooks

Kachigan, S.K. 1991. *Multivariate Statistical Analysis: A Conceptual Introduction*. New York: Radius.

Hutcheson, G.; and Sofroniou, N. 1999. *The Multivariate Social Scientist*. Thousand Oaks: Sage.

These textbooks are required. In addition, a number of readings are required, as indicated in the course schedule. Additional readings may occur in conjunction with specific assignments, for the purpose of explaining specific datasets and the questions they are intended to address.

## Readings

Agarwal, R., De, P., Sinha, A.P., and Tanniru, M. 2000. On the usability of OO representations. *Communications of the ACM*, 43.10, 83-89

Alemayehu, N. 2003. Analysis of performance variation using query expansion. *JASIST* 54.5, 379-391.

Boyack, K.W., and Börner, K. 2003. Indicator-assisted evaluation and funding of research: visualizing the influence of grants on the number and citation counts of research papers. *JASIST* 54.5, 444-461.

- Carroll, J.D. 1976. Spatial, non-spatial, and hybrid models for scaling. In *Psychometrika* 41, 439-463.
- Cooper, L.Z. 2002. Methodology for a project examining cognitive categories for library information in young children. *JASIST* 53.14, 1223-1231.
- Kruskal, J. 1977. The relationship between multidimensional scaling and clustering. In *Classification and Clustering*, 17-44. New York: Academic Press.
- Landauer, T. K.; P.W. Foltz; and D. Laham. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes*, 25:259-284.
- Lavie, T.; and N. Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of websites. *International Journal of Human-Computer Studies*, 60: 269-298.
- Papa, F., and Spedaletti, S. 2001. Broadband cellular radio telecommunication technologies in distance learning: a human factors field study. *Personal and Ubiquitous Computing* 5, 231-242.
- Powers, D.M.W. 1997. Unsupervised learning of linguistic structure: an empirical investigation. *International Journal of Corpus Linguistics*, 2.1, 91-131.
- Sun, A., Lim, E.-P., and Ng, W.-K. 2003. Performance measurement framework for hierarchical text classification. *JASIST* 54.11, 1014-1028.
- Tacq, J. 1997. Factor Analysis. Ch 9 of *Multivariate Analysis Techniques in Social Science Research*, 266-321. Thousand Oaks: Sage.
- Venables, W.N, and Ripley, B.D. 2002. Exploratory multivariate analysis. Ch 11 of *Modern Applied Statistics with S*, Fourth Edition, 301-330. Berlin: Springer Verlag.
- Zinnes, J.L., and Mackay, D.B. 1992. A probabilistic multidimensional scaling approach: properties and procedures. In F.G. Ashby, ed., *Multidimensional Models of Perception and Cognition*, 35-60. New York: Lawrence Erlbaum.
- Tversky, A., and Hutchinson, J.W. 1986. Nearest neighbor analysis of psychological spaces. *Psychological Review*, 93.1, 3-22.
- Watters, C., and Amoudi, G. 2003. GeoSearcher: location-based ranking of search engine results. *JASIST* 54.2, 140-151.

## Assignments:

The regular assignments are exercises with statistical techniques on datasets provided by the instructor. These assignments address the following topics:

- Final Paper/Project Due Finals Week. The final paper is the application of one or more multivariate

statistical techniques (generally one or two of the following: regression/ANOVA, Generalized Linear Modeling, Principal Components, Factor Analysis, Multi-Dimensional Scaling, Hierarchical Cluster Analysis) to a set of data selected by the student. Guidance will be provided on the selection of datasets and appropriate statistical techniques for analysis.

- Presentation Due Week 15. Each student will present her or his final paper project in the final week of class.

## Grading

Assignments in this class are evaluated according to the following table.

Assignment	value	number	Total
Assignments	5%	8	40%
Presentation	15%	1	15%
Final Paper	45%	1	45%

## Course Requirements

To receive a passing grade in this course, you must turn in all of the assignments and the term project and do your presentations. You cannot pass this course without doing all of the assigned work, however, turning in all of the work is not a guarantee that you will pass the course. All papers and assignments must be submitted on the dates specified in this syllabus. If you cannot submit an assignment or cannot deliver a presentation on the date it is due, it is your responsibility to discuss your situation with the instructor, in advance of the assigned date.

Your written, web-based, and/or oral work will be evaluated according to four criteria; it must:

1. Be clearly written, marked up, and/or presented, and checked for spelling and grammar;
2. Demonstrate a degree of insight into the concepts, issues, and trends in both the areas you investigate in the assignments and in the course content;
3. Demonstrate a degree of originality in your reviews, analyses and projects; and
4. Display familiarity with the appropriate literature.

Borderline grades will be decided (up or down) on the basis of class contributions and participation throughout the semester.

The following definitions of letter grades have been defined by student and faculty members of the Committee on Improvement of Instruction and have been approved by the faculty (November 11, 1996) as an aid in evaluation of academic performance and to assist students by giving them an understanding of the grading standards of the School of Library and Information Science:

### Grade GPA Meaning

A 4.0 Outstanding achievement. Student performance demonstrates full command of the course

materials and evinces a high level of originality and/or creativity that far surpasses course expectations

- A- 3.7 Excellent achievement. Student performance demonstrates thorough knowledge of the course materials and exceeds course expectations by completing all requirements in a superior manner
- B+ 3.3 Very good work. Student performance demonstrates above-average comprehension of the course materials and exceeds course expectations on all tasks as defined in the course syllabus
- B 3.0 Good work. Student performance meets designated course expectations, demonstrates understanding of the course materials and is at an acceptable level
- B- 2.7 Marginal work. Student performance demonstrates incomplete understanding of course materials.
- C+ 2.3 Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course materials
- C 2.0
- C- 1.7 Unacceptable work. Course work performed at this level will not count toward the MLS or MIS degree. For the course to count towards the degree, the student must repeat the course with a passing grade.
- D+ 1.3
- D 1.0
- D- 0.7
- F 0.0 Failing. Student may continue in program only with permission of the Dean.

Indiana University and School of Library and Information Science policies on academic dishonesty will be followed. Students found to be engaging in plagiarism, cheating, and other types of dishonesty can expect to receive an F for the course.

---

This page maintained by John C. Paolillo

---

Retrieved from <http://ella.slis.indiana.edu/~paolillo/Teaching/S604s08>  
Page last modified on January 15, 2008, at 10:02 AM