SAN DIEGO STATE UNIVERSITY
Graduate School of Public Health
Division of Epidemiology and Biostatistics

PH 628 Applications of Multivariate Statistics in Public Health 3 units
Fall 2006

Section	Day	Time	Location	Schedule No.
1	Mon Wed	2:00p – 3:15p	HH 122	25922

Instructor: John Alcaraz, Ph.D.	**Office location:** Hardy Tower 231
Office phone: (619) 594-1342	**Office hours:** Mon Wed 12:30p – 1:50p
E-mail: jalcaraz@mail.sdsu.edu	Wed 5:30p – 6:50p

Blackboard:
During the semester, course-related materials such as announcements, lecture notes, and homework solutions will be posted on Blackboard. Please check regularly for new materials.

Required texts:
– Kleinbaum, Kupper, Muller, Nizam: *Applied Regression Analysis and Multivariable Methods, 3rd edition.* [KKM]
– Afifi, Clark, May: *Computer-aided Multivariate Analysis, 4th edition.* [AC]
– Slymen: "PH 628: Applications of Multivariate Statistics in Public Health" (Customized Materials). [R]
– Slymen: "PH 628: Annotated SAS Output for Public Health" (Customized Materials). Note: The annotated output should be brought to *every* class meeting.

Grading System*:

Exercises: 25%	93 – 100 = A	73 – 77 = C
Project**: 35%	90 – 93 = A–	70 – 73 = C–
Final Exam: 40%	87 – 90 = B+	67 – 70 = D+
	83 – 87 = B	63 – 67 = D
	80 – 83 = B–	60 – 63 = D–
	77 – 80 = C+	0 – 60 = F

* All coursework will require using SAS on the PC.
** Paper describing an in-depth analysis you perform using one or more multivariate methods.

Dates for Coursework (subject to change):

	Date Assigned	Date Due
Exercise 1	Sep 13	Sep 27
Exercise 2	Sep 27	Oct 11
Exercise 3	Oct 4	Oct 18
Exercise 4	Oct 18	Nov 1
Exercise 5	Nov 1	Nov 15
Exercise 6	Nov 1	Nov 22
Project	Sep 20	Nov 29
Final Exam (open-book, open-notes)	Nov 15	Dec 6

Academic Ethics:
SDSU has a strict code of ethical conduct which students are expected to follow. See http://www.sa.sdsu.edu/srr/judicial/StudentConductCode.html for details. In particular, cheating will not be tolerated. You may not work together on the project or final exam, may not copy other students' work, and may not allow other students to copy your work. Anyone caught cheating will face disciplinary action.

Attendance: Although attending every class meeting is not required, it is strongly encouraged.

Prerequisites:
1) PH 627 or equivalent course work in multiple regression, analysis of variance and logistic regression.
2) Completion of the SAS computer class or equivalent knowledge of SAS.

Learning Objectives:
In this course, students will learn the appropriate use of multivariate methods for the analysis of health-related data with multiple dependent and independent measures where multivariate assessment and/or variable reduction are the primary goals. Students will become familiar with computer procedures in SAS commonly used in multivariate analyses. Using SAS, students will be able to perform the following statistical procedures:

1. Linear regression diagnostics. Students will be able to check for violations of the assumptions of multiple linear regression, to identify influential data points, and to check for collinearity.

2. Logistic regression diagnostics. Students will be able to assess goodness-of-fit of logistic regression models, to identify influential data points, and to check for collinearity.

3. Principal components analysis. Students will be able to construct a set of components which summarize the interrelationships among a set of variables. Students will be able to assess whether these components may be used in place of the original variables in other analyses.

4. Factor analysis. By constructing a set of factors, students will be able to verify whether hypothesized or expected interrelationships appear among a set of variables.

5. Cluster analysis. Students will be able to group together subjects (e.g., patients) according to similar values on measured variables, where such groupings are not specified in advance. This is primarily an exploratory technique.

6. Discriminant analysis. Students will be able to construct a rule based on a set of variables which optimizes the classification of subjects among two or more specified groups (e.g., with disease, without disease). Students will be able to assess the utility of the rule for classification.

7. Polychotomous logistic regression. Students will be able to test the association between a set of independent variables and a categorical dependent variable that has more than two unordered categories.

8. Ordinal regression. Students will be able to test the association between a set of independent variables and a categorical dependent variable that has more than two ordered categories.

9. Analysis of longitudinal data. Students will be able to analyze studies in which subjects are followed over time and repeated measurements of the outcome variables are taken on each subject.

10. Poisson regression. Students will be able to test the relationship between a set of independent variables and a dependent variable which counts the number of times a particular event occurs.

Additionally, students will learn to work independently or with minimal supervision to formulate and pursue an applied public health research question, and to communicate the results in writing.

Course Outline for PH 628 *

Topic	Related Book Chapters
1. Review of multiple linear & logistic regressions	
2. Regression diagnostics and goodness-of-fit in multiple linear and logistic regression a. Residual analysis b. Detecting outliers c. Detecting collinearity d. Goodness-of-fit statistics	KKM 12, R
3. Principal components analysis (PCA) a. Basic properties & geometric interpretation b. Using PCA to detect outliers and collinearity	AC 14, R
4. Factor analysis (briefly)	AC 15
5. Cluster analysis	AC 16
6. Discriminant analysis (DA) a. Basic properties b. Two-group DA c. DA for more than 2 groups d. Estimation of error rates & posterior probabilities e. Relationship to multivariate ANOVA	AC 11, R
7. Polychotomous logistic regression a. Basic properties b. Estimation and hypothesis testing c. Modeling and examples	R
8. Ordinal regression a. Basic properties b. Modeling and examples	R
9. Analysis of longitudinal data a. Introduction and examples b. Mixed effects models c. Modeling the covariance structure d. Special issues: missing data, attrition	R
10. Poisson regression (briefly)	R

* Lecture notes for all topics will be available on Blackboard.