THE UNIVERSITY OF CALGARY

Faculty of Science
Department of Mathematics & Statistics

Course Information Sheet

Course	STAT 525/625: Multivariate Analysis	**Session**	Winter 2005
Lecture	L01 MWF 10:00-10:50 am	**Room**	MS 427
Tutorial	T 10:00-10:50 am	**Room**	MS 571
Instructor	Alex R. de Leon		
Email	`adeleon@math.ucalgary.ca`		
Office	MS 554	**Phone**	220-6782
Office Hours	Whenever I'm in my office or by appointment		

Prerequisites STAT 421 or consent of the Division

NOTE: The Faculty of Science policy on pre- and co-requisite checking is outlined on page 199, of the 2004-2005 Calendar. **It is the students' responsibility to ensure that they have the prerequisite for the course, and if they do not they will be withdrawn from the course without further notice.**

1 **Fee Policy:** After the last day to drop/add courses (**January 21**, Friday), there will be no refund of tuition fees if a student withdraws from a course, courses or the session.

2 **Grading:** The University policy on grading and related matters is described on pages 43-52 of the 2004-2005 Calendar. In determining the overall grade in the course, the following weights will be used:

Assignments	[every 2 weeks]	30%
Mid-term Test	[1]	20
Project	[1]	10
Final Exam	[1]	40

There will be a final examination scheduled by the Registrar's Office. The use of aids such as open book, etc., is not permitted.

3 **Missed Components of Term Work:** The regulations of the Faculty of Science pertaining to this matter are outlined on page 200 of the 2004-2005 Calendar. It is the student's responsibility to familiarize himself/herself with these regulations.

4 **Academic misconduct** (cheating, plagiarism, or any other form) is a very serious offence that will be dealt with rigorously in all cases. A single offence may lead to disciplinary probation or suspension or expulsion. The Faculty of Science follows a *zero tolerance policy* regarding dishonesty. Please read the sections of the 2004-2005 University Calendar under the heading "Student Misconduct", pages 53-56.

5 **Required Text:**

Applied Multivariate Statistical Analysis (Fifth Edition)
by Johnson, R. A. & Wichern, D. W.
Prentice Hall, 2002.

6 **Assignments:** Assignments will include both data analysis and theoretical problems. They should be handed in before the beginning of the class, on the announced due date. **Each student should write up his/her assignment in his/her own words.**

7 **Tests:** Both the midterm and final tests are **closed-book-no-notes** examinations. The former is a 45-minute test to be written during a lecture class (see Table 1) while the latter is a 2-hour test to be scheduled by the Registrar's Office.

8 **General Comments:** This is a statistics course in which theory and applications are blended at a relatively high level. Background in linear algebra and regression analysis is of primary importance, whereas calculus and statistical inference are of secondary importance. Some helpful tips are as follows:

- Rewrite your notes in your own words **after each lecture** to make sure that you understand the material.
- Don't hesitate to see me right away if the lecture material is not clear to you.
- **On assignments:** Start working on your assignment early–do not expect to be able to finish it a few days before the due date. **Do not hand in your rough work!** If possible, rewrite your work at least once. Be careful to include a sufficient amount of explanation so I will not have to guess what you mean. I will want to see that you understand what you are writing, not merely that you arrive at the correct answer.

9 **Project:** The project will involve a comprehensive analysis of a multivariate data set chosen by the student. The report should be at most 10 type-written pages and should contain, in addition to the detailed analysis, a description of the scientific questions underlying the collection of the data, the reasons for the particular approach adopted, and an explanation of how the analysis addresses the questions involved. **Photocopies of published papers that describe the chosen data set should be included in the report as appendices.** Students are encouraged to try to find datasets related to their graduate research.

10 **Tutorials:** Some experience with the use of statistical packages will be helpful. Students may use whatever packages (e.g., Minitab, SPlus) they prefer. Once-a-week 50-minute lab tutorials have been booked for the course in MS 571. You may use these to work on the assignments. Certain sessions will be devoted to special topics.

11 **Dates to remember:**

January 18, 2005	*First tutorial session*
February 18, 2005	*Mid-Term Test*
February 20-27, 2005	*Reading Week (No Classes)*
April 12, 2005	*Last tutorial session*
April 15, 2005	*Last lecture class*
April 15, 2005	*Last day to withdraw from Winter Session classes*

12 **Grading System:** At the end of the term, a summary score will be computed from assignment and test marks which will be used to rank everyone in order of merit. I will then decide whether the class as a whole is average, above average or below average, which will determine how many students should fall into each of the possible letter grades. The grades will be assigned accordingly and they will reflect my judgment and assessment of both absolute achievement and relative performance in the class. **No fixed scale will be used to convert your end-of-term summary scores to letter grades.**

13 **Course Webpage:** Assignments and other relevant supplementary materials may be downloaded from the course blackboard. Solutions to the assignments will be posted on the course blackboard, when possible.

Table 1: *Tentative STAT 525/625 L01 Course Outline*

Lecture	Date	Readings	Topics
1	M Jan 10	Ch. 1	Preliminaries
2	W Jan 12	Ch. 2 & 2A	Vector geometry, Cauch-Schwarz, eigenvalues/vectors
3	F Jan 14	Ch. 2 & 2A	Spectral decomposition, square roots, quadratic forms
4	M Jan 17	Ch. 2 & 2A	Random matrices, moments, partitioning, trace
5	W Jan 19	Ch. 3	Sample mean/covariance, linear combinations
6	F Jan 21	Ch. 3	Generalized & total variance, principal components
7	M Jan 24	§ 4.1-4.2	Definition of multivariate normality
8	W Jan 26	§ 4.1-4.2	Basic properties of MVN dist'n, conditional dist'ns
9	F Jan 28	§ 4.3	MLE of parameters of MVN dist'n
10	M Jan 31	§ 4.4-4.5	Joint distribution of $\overline{\mathbf{x}}$ & $\mathbf{S}$
11	W Feb 2	§ 4.6-4.8	Assessing normality
12	F Feb 4	§ 4.6-4.8	Transformations
13	M Feb 7	§ 5.1-5.3	T^2 test for $\boldsymbol{\mu}$
14	W Feb 9	§ 5.1-5.3	T^2 test as LR or UI procedure
15	F Feb 11	§ 5.4-5.5	Confidence regions, simultaneous comparisons
16	M Feb 14	§ 6.3	Two-sample problem
17	W Feb 16	§ 6.1-6.3	Paired comparisons, repeated measures
	F Feb 18	**Mid-Term Test**	
Feb 20-27: READING WEEK			
18	M Feb 28	§ 6.4	One-way MANOVA
19	W Mar 2	§ 6.5	Simultaneous confidence intervals, example
20	F Mar 4	§ 6.6	Two-way MANOVA
21	M Mar 7	§ 6.7	Profile analysis
22	W Mar 9	§ 7.1-7.7	Multivariate regression: Estimation
23	F Mar 11	§ 7.1-7.7	Multivariate regression: Testing
24	M Mar 14	§ 7.1-7.7	Multivariate regression: Prediction; example
25	W Mar 16	§ 7.8-7.9	Random regressors
26	F Mar 18	§ 8.1-8.2	Principal components
27	M Mar 21	§ 8.3-8.5	Principal components
28	W Mar 23		PCA example
29	M Mar 28	§ 11.1-11.2	Separation & classification
30	W Mar 30	§ 11.2-11.3	Two MVN populations
31	F Apr 1	§ 11.4	Example; evaluating classification functions
32	M Apr 4	§ 11.4-11.5	Error rates, Fisher's discriminant function
33	W Apr 6	§ 11.6-11.7	Classification & discrimination for > 2 pop'ns
34	F Apr 8	§ 12.1-12.6	Clustering
35	M Apr 11	§ 12.1-12.6	Clustering
36	W Apr 13	§ 12.1-12.6	Clustering
37	F Apr 15	§ 12.1-12.6	Clustering

STATISTICS 525/625
Guidelines for the Project

The project consists in finding a suitable data set and analyzing it using multivariate methods. The data set should not be taken from a textbook. Possible sources for data sets include any subject-area journals (e.g., engineering, medicine, biology, etc.), a Data Library, or the web. Better yet, you may want to collect your own data. Each student will give an oral presentation and prepare a written report. The organization of the report should generally follow the following:

1. **Introduction.** Describe the source of the data, how and why the data were originally collected, the definition and interpretation of the variables, and the question/s your analysis intends to answer. Briefly summarize your approach in addressing the questions.

2. **Initial examination of the data.** Describe the results of the preliminary examination of the data. Were there any outliers? How did you deal with them? Was a transformation of the data needed?

3. **Analysis.** This typically is the longest part of the report. This may involve formal statistical inference, such as fitting a model and testing hypotheses, or it may entail more exploratory methods, such as cluster analysis. Make sure that you state clearly the assumptions underlying your analysis, check the plausibility of these assumptions (e.g., by residual analysis), and if possible, relate your assumptions to the method of data collection. Your analysis should be supported by appropriate graphical and/or tabular displays.

4. **Conclusion.** Summarize your results. Did the data provide clear answers to your questions or are more data needed? Did your analysis raise new questions?

The project will be graded on clarity of exposition as well as content. The introduction should not contain too many statistical terminologies. Pretend you are a consultant preparing a report for a scientist who has no strong statistical background outside of an introductory course taken several years ago. The analysis section should be more technical, but the emphasis should still be on the relevance of the methodology to the study and not the calculations. The report should be at most 10 pages, including graphs and tables. Computer output should be integrated into the report and graphs/tables should be clearly labeled. You must append photocopies of any paper where the data set appears to your report as appendices.

Oral Presentation: April 18, 2005
Written Report : On or before April 22, 2005

Reference: Chatfield, C. (1988). *Problem Solving: A Statistician's Guide*, Chapman & Hall.