STATISTICS AND DATA ANALYSIS IN GEOLOGY

MWF 10:30 – 11:30, 136 Natural Science 3 credits

Overview:
The purpose of this course is to give students a working knowledge of the various types of statistical tests used in geology and related sciences. Through computer-based exercises the students will explore the uses and pitfalls of statistical techniques and when certain tests should be done and how to interpret results of tests. Students should have some experience with personal computers and have used word processors and spreadsheet programs. GEOS 225 and STAT 200 (or equivalents) are prerequisites for the course. Math 200 is recommended. Let me know if you do not have a strong math background.

Instructor: Paul Layer, 368 Natural Science Office phone 474-5514
player@gi.alaska.edu

Office hours: Briefly after class and Monday, Tuesday, Wednesday, Thursday, 1:00 – 2:00, and Monday 3:30 – 5:00 (for computer help)

Required Text: J.C. Davis, Statistics and Data Analysis in Geology, 3rd edition (not 2nd or 1st!!!)

****

General outline

WEEK	TOPIC	CHAPTER
1-5	Univariate Statistics	Chapter 1, 2
6	Linear Algebra	Chapter 3
7	Regression	Chapter 4
	SPRING BREAK	
8	Analysis of Circular Data	Chapter 5
9	Time Series	Chapter 4
10-11	Map Analysis	Chapter 5
12-14	Multivariate Analysis	Chapter 6

****

GRADING: 55% of the grade will be based on homework assignments. The midterm exam is worth 20%, the final exam is worth 20%. 5% of the grade is for class participation/attendance.

LATE POLICY: All homework assignments are due at the BEGINNING OF CLASS of the due date (therefore, skipping class to work on an assignment will not gain you anything). If one class period late — 5% off; 2 class periods late — 10% off; one week late — 15% off; greater than 1 week — 30% off. Assignments from the first half of the semester must be submitted before 5:00 PM Friday March 10 or they will not be graded. (Deadlines are flexible in extenuating circumstances, travel, computer crashes, and by prior arrangement only). If you are planning to miss a class or two, please let me know in advance.

I expect students to follow the Student Code of Conduct (pages 73-74 of the 2005-2006 UAF Catalog).

HOMEWORK ASSIGNMENTS: These will be a combination of computer and non-computer exercises. For the non-computer questions, these will require calculations and graphing and may require the use of computer programs like Excel or SPSS. For those more computer oriented, the emphasis will be on that you can produce the necessary chart and table, present it in an interpretable format and write a short paragraph explaining how to interpret the statistical information. **Assignments distributed on Wednesday will be due on the following Monday, assignments distributed on Friday will be due the following Wednesday**, so there is one class period for questions. IF YOU PUT THESE OFF UNTIL THE NIGHT BEFORE, YOU WILL NOT BE ABLE TO FINISH THEM!! If more than 10 students are in the class, it might get crowded near the due date. Plan ahead. Please feel free to help each other with computer or calculation-related questions, but any written work should be your own. Show your work, not just your answers. If you do calculations using Excel, indicate the formulae used so I can figure out how you got your answer. For some of the exercises, you might have to use statistical programs in the Geology computer lab, and each of you will be given an account in the lab.

MIDTERM EXAM: March 8, IN-CLASS, CLOSED BOOK. This test will include definitions of terms, short answers, and basic problem solving. You will need a calculator (a basic one for addition, multiplication, etc). The test will cover through Lecture 19. I will distribute review sheets.

FINAL EXAM: The final exam will be essay/short answer questions and based on concepts presented in the weekly lecture material and assignments, and will be an IN-CLASS, OPEN BOOK exam. The emphasis of this exam will be qualitative, not quantitative and cover the entire semester. That is, I will test your understanding of the uses (applicability, limitations) of the various statistical tests and will not be computational. The Final examination for this class is scheduled for Monday May 8. This doesn't give much time after the last class. We can discuss rescheduling or a take-home final instead.

All outstanding homework assignments will be due by Thursday, May 11.

If you are having problems or just feel uncomfortable with the computer or programs, or your statistical background is a bit weak, please see me. I have reserved time in the computer lab on Monday afternoon for a help session. If you have a documented disability that requires additional time on homework assignments or tests, or if you require other accommodation, please let me know within the first two weeks of the semester. If you have questions regarding a homework exercise, or feel that you are falling hopelessly behind, please see, call, or e-mail me ASAP.

Attached is a tentative class schedule. What is actually covered on any particular day may change as the semester progresses. Lecture notes for each class are available on a CD as .PDF files. These may also be available in the computer lab. **Before class, please read the lecture notes and relevant chapters or sections in Davis**. The CD will also contain data sets discussed in Davis and a large data set that we will use with some of the homework exercises (LPMP.xls).

My promise to you: If you turn in all of the assignments with reasonable effort and in a timely manner, if you put in a sincere effort on the exams and you attend and participate in class, you will pass the course (C). Although I tend to look at the performance of the class as a whole and do not have hard-and-fast grade cut-offs, in recent years, with roughly the present homework, midterm exam and final exam weighting, my final grading has run: 90's == A, 80's == B, 70's == C, 60's == D.

Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
0:	1	Jan	20	Introduction, course overview, variance	1-9	Review syllabus, CD, lecture notes
1:	2		23	Types of data, precision, Exploratory data analysis: Stem + leaf, Letter value	25-33	
	3		25	Box plots Univariate data distributions	97-102	Homework 1 (EDA, box plots)
	4		27	Univariate statistics	34-39	Homework 2 (Univariate statistics)
2:	5 and LPMP.pdf		30	Probability Introduction to the data set	11-24	
	6	Feb	1	Normal distribution, probability plots, Central Limits Theorem	58-60	Homework 3 (Lime Peak Data Set)
	7		3	Hypothesis testing: Z-test	55-58, 60-68	Homework 4 (Hypotheses and CLT)
3:	8		6	Hypothesis testing: t-test	68-74	
	9		8	2-sample t-test, F-test	75-78	Homework 5 (Hypothesis tests)
	10		10	Chi-Square test, testing data distributions, normalcy tests	92-96, 107-112	Homework 6 (Lime Peak EDA)
4:	11		13	The Poisson distribution Data distribution on maps Non-parametric statistics	19, 102-112, 184-185, 299-312	
	12		15	Propagation of errors	Only notes	Homework 7 (Error propagation, Chi-square)
	13		17	Analysis of variance	78-92	Homework 8 (Non-parametric stats)
5:	14		20	Post-hoc tests, non-parametric ANOVA	105	
	15		22	Joint probability distribution, covariance, correlation	40-46, 74-75, 105-107	Homework 9 (ANOVA)
	16		24	Matrix algebra	123-131	Homework 10 (ANOVA, correlation)

Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
6:	17	Feb	27	Gauss-Jordan method, matrix inversion, inverses and determinants	132-140	
	18	Mar	1	Eigenvalues and eigenvectors	141-153	Homework 11 (Matrices)
	19		3	Linear regression	191-207	Homework 12 (ANOVA and t-tests on big data set)
7:	20		6	Non-linear regression, Midterm review	207-214	
			8	**MIDTERM EXAM**		
	21		10	Residuals, structural regression, error limits for regression	214-220, 227-228	
				SPRING BREAK		
8:	22		20	Weighted regressions, forced regressions, robustness,	220-227	
	23		22	Circles and spheres	316-330	Homework 13 (Regression)
	24		24	Circles and spheres, Other problems with geologic data	331-342 46-50	Homework 14 (Circles)
9:	25		27	Time Series, autocorrelation	159-191, 243-254	
	26		29	Time Series, harmonic analysis, fractals	266-277, 342-354	Homework 15 (Time series)
	27		31	Computer contouring, point interpolation, trend surface analysis (TSA)	293-299 370-416	Homework 16 (Regression on the big data set)
10:	28	Apr	3	The semivariogram	254-265	
	29		5	Kriging	416-428	Homework 17 (the semivariogram)
	30		7	Universal Kriging	428-443	Homework 18 (Kriging and Contouring)

Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
11:	31	Apr	10	Analysis of multivariate data, presentation of multivariate data	461-470, 479-487	
	32		12	Multivariate regression, MANOVA		Homework 19 (Point interpolation)
	33		14	Multivariate discriminant analysis (MDA)	471-479	Homework 20 (Multivariate regression)
12:	34		17	MDA, continued, Clusters	572-577	
	35		19	Cluster analysis	487-500	Homework 21 (MDA)
	36		21	Comparison of cluster methods and MDA		Homework 22 (Clusters 1)
13:	37		24	Principal component analysis (PCA)	500-525	
			26	Principal component analysis and R-mode factor analysis	526-540	Homework 23 (Clusters 2)
	38		28	**NO CLASS**		***All Campus Day***
14:	39	May	1	R-mode factor analysis		
	40		3	Q-mode factor analysis and Canonical Correlation	540-571, 577-584	Homework 24 (Factor analysis)
	41 (no notes)		5	Wrap-up and review		
			8	**FINAL EXAM**		Monday