STATISTICS AND DATA ANALYSIS IN GEOLOGY

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

Instructor: Franz Meyer; Office: 106D, Westridge Research Building
Phone: 474-7767 fmeyer@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, if needed)

Course description: GEOS 430 is a required course for undergraduate Geoscience majors (and a recommended one for graduate students in a number of fields) that introduces students to applications of statistical data analysis including computer-supported geologic applications of elementary statistics, time-series analysis, trend-surface analysis, factor analysis, cluster analysis, discriminant analysis and multiple regression. It draws on previous geoscience or other science experience for examples of how statistics can be applied to complex data sets.

Prerequisites: GEOS 225 and STAT 200X: MATH 200 is recommended. Let me know if you do not have a strong math background or if you have not had these courses. Students should have some experience with personal computers and have used word processors and spreadsheet programs.

Required Text: J.C. Davis, Statistics and Data Analysis in Geology, 3rd edition (not 2nd or 1st!!!). There will be a CD distributed with lecture notes that will be discussed in class.

Course Goals and Student Learning Outcomes: 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 will be evaluated on how well they can recognize and apply statistical tests to a variety of data sets and problems.

General outline

WEEK	TOPIC	CHAPTER
1	Introduction and overview	Chapter 1
2-6	Univariate Statistics	Chapter 2
7	Linear Algebra	Chapter 3
	SPRING BREAK	
8	Regression	Chapter 4
9	Analysis of Circular Data	Chapter 5
10	Time Series	Chapter 4
11	Map Analysis	Chapter 5
12-15	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. 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. This year we are implementing + and – grades, which will be used to better evaluate borderline cases.

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).

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 through Homework 11 must be submitted before 5:00 PM Friday March 7 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.
All outstanding homework assignments will be due by Friday, May 7.

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 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, the computer lab 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 22, 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 20. 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 the test will not be computational. The Final examination for this class is scheduled for Monday May 10.

Getting Help: If you are having problems or just feel uncomfortable with the computer or programs, or if 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.

Disabilities Services: The Office of Disability Services implements the Americans with Disabilities Act (ADA), and insures that UAF students have equal access to the campus and course materials. I will work with the Office of Disabilities Services (203 WHIT, 474-7043) in order to provide reasonable accommodation to students with disabilities.

I expect students to follow the Student Code of Conduct (pg. 47-48 of the 2009-2010 UAF Catalog).

Below is a 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).

Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
1:	1	Jan	22	Introduction, course overview, variance	Pages 1-9	Review syllabus, CD, lecture notes
2:	2		25	Types of data, precision, Exploratory data analysis: Stem + leaf, Letter value	33	
	3 and LPMP.pdf		27	Box plots Introduction to the data set		Homework 1 (EDA, box plots)
	4		29	Univariate data distributions Univariate statistics – central tendency	24-35, 97-102	Homework 2 (Lime Peak Data Set)
3:	5	Feb	1	Univariate stats – spread, skew, shape, Probability	11-24, 34-39	
	6		3	Normal distribution, probability plots, Central Limits Theorem	58-60	Homework 3 (Univariate statistics)
	7		5	Hypothesis testing: Z-test	55-58, 60-68	Homework 4 (Lime Peak EDA)
4:	8		8	Hypothesis testing: t-test	68-74	
	9		10	2-sample t-test, F-test	75-78	Homework 5 (Hypotheses and CLT)
	10		12	Chi-Square test	92-96	

Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
5:	11	Feb	15	Propagation of errors		
	12		17	Non-parametric statistics The Poisson distribution	102-112, 184-185, 19	Homework 6 (Error propagation)
	13		19	Testing data distributions, normalcy tests, data distribution on maps	107-112, 299-312	Homework 7 (Non-parametric stats)
6:	14		22	Analysis of variance	78-92	
	15		24	Post-hoc tests, non-parametric ANOVA	105	Homework 8 (Chi-square, ANOVA)
	16		26	Joint probability distribution, covariance, correlation, closed data sets	40-50, 74-75, 105-107	
7:	17	Mar	1	Matrix algebra	123-131	
	18		3	Gauss-Jordan method, matrix inversion, inverses and determinants	132-140	Homework 9 (Correlation)
	19		5	Eigenvalues and eigenvectors	141-153	Homework 10 (Matrices)
				SPRING BREAK		
8:	20		15	Linear regression	191-207	
	21		17	Non-linear regression	207-214	Homework 11 (ANOVA, big dataset)
	22		19	Residuals, Structural regression, error limits for regression	214-220, 227-228	
9:			22	**MIDTERM EXAM**		
	23		24	Forced regressions, weighted regressions, Robustness, statistics on circles	220-227	Homework 12 (Regression)
	24		26	Statistics on circles and spheres (continued)	316-342	

Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
10:	25	Mar	29	Time Series, autocorrelation, cross correlation	159-191, 243-254	
	26		31	Time Series, harmonic analysis, fractals	266-277, 342-354	Homework 13 (Time series)
	27	Apr	2	Computer contouring, point interpolation, trend surface analysis (TSA)	293-299 370-416	
11:	28		5	The semivariogram	254-265	
	29		7	Kriging	416-428	Homework 14 (Regression on the big data set)
	30		9	Universal Kriging	428-443	
12:	31		12	Analysis of multivariate data, presentation of multivariate data	461	
	32		14	Multivariate regression, MANOVA	462-470, 479-487	Homework 15 (Multivariate regression)
			16	Multivariate discriminant analysis (MDA)	471-479 572-577	
13:	33		19	Cluster Analysis	487-500	Homework 16 (MDA)
	34		21	Comparison of cluster methods and MDA		
	35		23	**NO CLASS**		***UAF SpringFest***
14:	36		26	Principal component analysis (PCA)	500-525	
	37		28	Factor Analysis	526-540	Homework 17 (Clusters & Factors)
	38		30	Principal component analysis and R-mode factor analysis case studies		
15:	39	May	3	Q-mode factor analysis and Canonical Correlation	540-571, 577-584	
			10	**FINAL EXAM**		Monday 10:15 – 12:15