STATISTICS AND DATA ANALYSIS IN GEOLOGY

MWF 11:45 – 12:45, 202 Reichardt   3 credits

Instructor: Franz Meyer; Office: 106D, Westridge Research Building
                              Phone: 474-7767   fjmeyer@alaska.edu
                   Office hours: Briefly after class and Monday, Wednesday, Thursday, 2:00pm - 3:00pm

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 the basics and 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 exercises and mini projects the students will explore the uses and pitfalls of statistical techniques and learn 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**	**Block**
1 – 2	**INTRODUCTION AND OVERVIEW**	
2 – 6	**UNIVARIATE STATISTICS**	**BLOCK 1**
7	PROJECT WORK 1: "UNIVARIATE STATISTICS"	
8 – 12	**MATHEMATICAL BASICS OF BIVARIATE AND MULTIVARIATE STATISTICS**	**BLOCK 2**
10	**SPRING BREAK**	
11	**MIDTERM EXAM**	**March 28**
12 – 16	**GEOSTATISTICS AND MULITVARIATE STATISTICS**	**BLOCK 3**
13	PROJECT WORK 2: "EXAMPLES OF GEOSTATISTICS"	
16	PROJECT WORK 3: "CLUSTERING AND OTHER MULTIVARIATE ANALYSIS METHODS"	
17	**FINAL EXAM**	**May 9**

Grading: 55% of the grade will be based on homework assignments (including the write-up of project work). 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, on average, the final grading has run: 90's == A, 80's == B, 70's == C, 60's == D. The class provides + and – grades.

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 that is specified on the homework sheets (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. (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 9.

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. 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 28, 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 21. 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 Friday May 9.

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 (208 WHIT, 474-5655) in order to provide reasonable accommodation to students with disabilities.

I expect students to follow the Student Code of Conduct (pg. 52 of the 2013-2014 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. **Before class, please read the lecture notes and relevant chapters or sections in Davis**. The CD will also contain data sets required for homework exercises.

	Week	Lecture # (readings)	Date		Topic	Readings in Davis	Assignment distributed
Basics about Data Analysis	1:	1	Jan	17	Introduction, course overview		Review syllabus
	2:			20	**MARTIN LUTHER KING DAY**		**NO CLASS**
		2		22	Data Types, Definitions, and Exploratory Data Analysis		
		3		24	Visualization of Data – Box Plots and other biased and unbiased plotting types		Homework 1 (Boxplots)

BLOCK 1: UNIVARIATE STATISTICS							
Basics about Univariate Statistics	3:	4		27	Univariate data distributions and parameters that describe	34-39	
		5		29	Normal Distribution: The mother of all statistical data distributions		
		6		31	Non-Normal things and the Central Limits Theorem	58-60	Homework 2 (Univariate statistics)
Hypothesis Testing	4:	7	Feb	3	Hypothesis testing theory; Z-test	55-68	
		8		5	The t-test	68-75	
		9		7	2-sample t-tests and F-tests	75-78	Homework 3 (Hypothesis testing)
	5:	10		10	The Chi-Square test – transition to non-parametric statistics	92-96	
Non-parametric statistics		11		12	The basics of non-parametric statistics and Non-parametric hypothesis testing	102-107	
		12		14	Applications of non-parametric statistics: Testing spatial data distributions; normalcy tests	107-112, 184-185, 299-312	Homework 4 (Non-parametric stats)

ANOVA	6:	13		17	Analysis of Variance: a generalized procedure for comparing datasets	78-92	
		14		19	Post-Hoc tests & non-parametric ANOVA		
		15		21	Joint probability distribution, covariance, correlation	40-50, 105-107	Homework 5 (Project 1)

Project Work	7:	Feb 24 - 28	<u>**Project Work 1:**</u> **Applying Univariate Statistics to the Real World**

BLOCK 2: SOME MATHEMATICAL BASICS OF BIVARIATE AND MULTIVARIATE STATISTICS							
Finding patterns in multivariate data	8:	16	Mar	3	Matrix algebra: A mathematical tool for solving n-dim. linear equation systems	123-131	
		17		5	Matrix inversion: Solving linear equation systems using matrix math	132-140	
		18		7	Eigenvalues and eigenvectors	141-153	Homework 6 (Correlation & Matrix math)
Regression Analysis	9:	19		10	Least-squares regression methods	191-214	
		20		12	Reduced major axis and structural regression techniques	214-220	
		21		14	Forced and weighted regressions	220-227	
	Week 10 (Mar 17 – 21): SPRING BREAK						
	11:	22		24	Error Propagation and Monte Carlo Simulation	227-228	
Sequenced Data		23		26	Analyzing sequence data: Autocorrelation & cross-correlation	243-254	Homework 7 (Regression and Time Series analysis)
				28	**MIDTERM EXAM**		
	12:	24		31	Finding patterns in sequence data: Harmonic analysis and fractals	266-277, 342-351	

					BLOCK 3: GEOSTATISTICS AND MULITVARIATE STATISTICS		
Spatial Surface Estimation		25	Apr	2	Computer contouring, interpolation, trend surface analysis (TSA)	293-299 370-416	
		26		4	The semivariogram	254-265	
	13:	27		7	Kriging: A geostatistical method for data interpolation and approximation	416-443	Homework 8 (Project 2)
Project Work		Apr 9 - 11			**Project Work 2:** **A Two Day Geostatistics Experiment**		
Analysis of Multivariate Data	14:	28		14	Discussion on categories of multivariate analysis techniques	461	
		29		16	Multivariate regression and multivariate ANOVA (MANOVA)	462-470 479-487	
		30		18	Multivariate Discriminant Analysis	471-479 572-577	
	15:	31		21	Cluster Analysis	487-500	
		32		23	Principal component analysis (PCA) and Factor Analysis I	500-525	
				25	**UAF SPRINGFEST**		**NO CLASS**
	16:	33		28	PCA and Factor Analysis II	526-540	Homework 9 (Project 3)
Project Work		Apr 30 – May 02			**Project Work 3:** **Data Clustering: Comparing MDA, Cluster Analysis, and PCA**		
			May	5	Semester Summary; Discussions		**Last day of Instruction**
				9	**FINAL EXAM**		Friday 10:15 – 12:15