

Course Announcement Fall 2004, MATH 583: Advanced Topics in Statistics:
Applied Robust Statistics
MWF 9:00-9:50 EGRA 0320

Statistics is the science of extracting useful information from data. Often there is a set of assumptions that makes up a model that is hopefully a useful approximation to the data. *Robust statistics* tries to give useful results when a specific assumption is violated. Course notes will be available from Kopies and More (809 S. Illinois Ave.) by Aug. 16.

One assumption is that the data observations are iid from a known distribution such as the normal distribution. Often this assumption is violated because outliers, cases that lie far away from the bulk of the data, are present. *Outliers* could be due to recording errors, eg 1000 pound women, 6 inch tall men, etc. *Distributionally robust statistics* can give useful results when outliers are present.

Regression is the study of the conditional distribution of the response variable Y given a vector of predictors \mathbf{x} . If the form of the regression model is unknown or misspecified, then *regression graphics* can be used to suggest a model or to detect the error.

This course covers distributionally robust statistics and regression graphics. *Robust statistics can be used to make many of the most used statistical procedures better.* The course notes show how to improve multiple linear regression, multivariate models, and models for categorical data analysis (3 of the most important applied methods in statistics). The *single index model*, widely used in econometrics, is also discussed. At the University of Illinois, this type of course is taken by *graduates in math, statistics, econometrics and electrical and computer engineering.* Additional topics are listed below.

- A) Outlier detection in the location model with plots and the sample median.
- B) Multiple linear regression, especially response transformations, variable selection, goodness of fit and lack of fit plots, and outlier detection.
- C) Robust estimators of multivariate location and dispersion with applications including outlier detection and a plot for determining whether the data is from a multivariate normal distribution or some other elliptically contoured distribution.
- D) Regression graphics procedures for 1D regression models (including single index models) with emphasis on exploratory data analysis, goodness of fit plots, and variable selection.
- E) Generalized linear models focusing on logistic regression and loglinear regression. Emphasis is on goodness of fit and lack of fit plots and variable selection.

Overlap with previous classes: material from chapters 5 and 6 of the course notes was used in my Math 484 class on multiple linear regression and design. Material from chapter 13 was used in my Math 485 class on categorical data analysis. Material from chapters 5, 6, 10, 11 and 12 was used in my 2000 Math 583 class on regression graphics.

The *prerequisites* for this class are Math 221 and Math 483 or Math 480. You should be familiar with the normal, gamma, binomial, Poisson and exponential distributions, confidence intervals and hypothesis testing. *Heavy use of the computer* will be made.

For more information contact David Olive, 261 Neckers.
Phone: 453-6566, Email: dolive@math.siu.edu

Math 583, Section 001, Fall 2004. MWF 9:00-9:50 EGRA 0320

Instructor: David Olive

Text: *Applied Robust Statistics* by David Olive available from Kopies and More (809 S. Illinois Ave.).

Office: (J.W.) Neckers 261 (wing A) *Phone:* (618)-453-6566

email: dolive@math.siu.edu

Office hours: MWThF 1:20-2:35, TH 12:20-1:20

I am also available by appointment and on a walkin basis, especially before and after class.

This course covers robust statistics including outlier detection and procedures from regression graphics. Variable selection and plots for goodness and lack of fit will be important.

Final (emphasis is on the above topics) Wed. Dec. 15 7:50–9:50AM (early in the morning), but the final may be replaced by a class project.

The grading and schedule below are tentative. (Drop day is Monday, October, 18.)

Except for the last week of classes, 2 homeworks may be turned in one class period late (ie on Monday) with no penalty. A third late will be accepted with 25% penalty. One or more sheets of notes will be allowed on quizzes and exams. A calculator is permitted.

Grading:

HW	400		Quizzes	100	
exam1	100	exam 2	100	exam 3	100
final	200	or project		total	1000
min. grade	points	min. grade	points	min. grade	points
A	900-1000	B	800-899	C	700-799
D	550-699				

Week of	MON	WED	FRI
Aug 23	Intro	2.1	2.2,2.3
Aug 30	lab	2.4, Q1	4.6, HW1
Sept 6	no class	5.1, Q2	5.1, HW2
Sept 13	5.2	5.3, Q3	6.1, HW3
Sept 20	lab	6.2,	Exam 1
Sept 27	6.4	6.4, Q4	7.6 HW4
Oct 4	lab	8.2, Q5	10.1, HW5
Oct 11	10.2	10.2	Exam 2
Oct 18	10.3	10.5, Q6	11.1, HW6
Oct 25	lab	11.2, Q7	12.1, HW7
Nov 1	12.2	12.2, Q8	12.3, HW8
Nov 8	12.3	12.4, Q9	12.4, HW9
Nov 15	lab	13.2, Q10	13.2, HW 10
Nov 22	no class	no class	no class
Nov 29	13.4	13.4, Q11	13.5 HW11
Dec 6	13.6	Exam 3	13.6