Course Announcement Fall 2020, MATH 583: Advanced Topics in Statistics: Robust Statistics MWF 10:00-10:50 Lawson 0201.

Students may take Math 583 up to 4 times (12 hours).

*Instructor:* David Olive

Text: *Applied Robust Statistics* (2008) at (http://parker.ad.siu.edu/Olive/ol-bookp.htm) and *Robust Statistics* (2020) (http://parker.ad.siu.edu/Olive/robbook.html) both by David Olive.

**Course Webpage:** http://parker.ad.siu.edu/Olive/M583.html

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

*email:* dolive@siu.edu   *Office hours:* MWThF 9-9:30, 12-1.

I am also available by appointment and on a walkin basis, especially before and after class. The course webpage will make some things available on the internet for students with covid concerns.

The *prerequisites* for this class are a Calculus based introduction to Statistics (Math 483, prereq Math 250) and familiarity with matrices and vectors (e.g., Linear Algebra: Math 221). You should be familiar with vectors, matrices, the normal, gamma, binomial, Poisson, and exponential distributions, confidence intervals, and hypothesis testing. A course in regression (Math 484, Poisson regression and logistic regression from Math 485, Cox regression from Math 474) would be useful. The $R$ software is used.

This course covers Robust Statistics with interest in outlier resistance and in methods with good large sample theory. I used *Applied Robust Statistics* for this course, but put some of the material in the Math 586 text Olive, D.J. (2017), *Robust Multivariate Analysis*, Springer, New York, NY. *Robust Statistics* is for the new course, but is being written this semester. 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, models for Statistical Learning, and models for categorical data analysis (4 of the most important applied methods in statistics). Additional topics are listed below.

A) The location model: outlier detection with plots and the sample median, prediction intervals with the shorth.

B) Multiple linear regression, especially response transformations, variable selection, goodness of fit and lack of fit plots, and outlier detection. Methods include least squares and Statistical Learning alternatives such as lasso.

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) Other regression models: 1D regression models with emphasis on exploratory data analysis, goodness of fit plots, and variable selection. Generalized linear models focusing on logistic regression and Poisson regression. Maybe AR($p$) time series. Emphasis is on goodness of fit and lack of fit plots and variable selection.

Overlap with previous classes: material from chapters 5 and 7 of the 2008 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. Some of the material is used in my classes on Statistical Learning and Multivariate Analysis.

The grading and schedule below are tentative. (Drop day is Sunday, October, 25, but paperwork is needed by the Oct. 23.) I sometimes give a $B+$ and $C+$.

Students receive a WF if they stop attending class and fail. An INC is given if for reasons beyond their control, students engaged in *passing* work are unable to complete all class assignments. Except for the last week of classes, 2 homeworks may be turned in one class period late (i.e. 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.

(Cumulative) Final: Monday, Dec. 7, 8-10AM. **Or do a project due during finals week.**

*Grading:*

| HW | 300 | | Quizzes | 100 | |
|---|---|---|---|---|---|
| exam1 | 100 | exam 2 | 100 | exam 3 | 100 |
| final | 300 | 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 17 | Intro | 2.1 | 2.2,2.3,2.4 |
| Aug 24 | 2.4,2.5 | 2.5,2.6, Q1 | 2.6,2.14, HW1 |
| Aug 31 | 2.14, 3.1 | 3.1,3.2, Q2 | 3.2,3.3, HW2 |
| Sept 7 | no class | 3.3,3.4, Q3 | 3.4, 3.6, HW3 |
| Sept 14 | 3.6 | 3.6 | Exam 1 |
| Sept 21 | 3.7,3.8, | 3.10,3.11, Q4 | 4.1, HW4 |
| Sept 28 | 4.2 | 4.2, Q5 | 4.3, HW5 |
| Oct 5 | 5.3,5.4 | 5.6,5.7, Q6 | 5.7,5.8,5.9, HW6 |
| Oct 12 | 5.9, 6.1, | 6.2, Q7 | 7.1, HW7 |
| Oct 19 | 7.1 | 7.1,7.2 | Exam 2 |
| Oct 26 | 7.3, | 7.3, Q8 | 7.4, HW8 |
| Nov 2 | 7.4,7.5,7.6, | 7.8,7.9, Q9 | 7.11, 7.13,8.1, HW9 |
| Nov 9 | 9.1,9.2 | no class | 9.2, Q10, HW 10 |
| Nov 16 | 10.3 | 10.4, Q11 | 10.5, HW11 |
| Nov 23 | no class | no class | no class |
| Nov 30 | Exam 3 | 10.8 | rev |