Course Announcement Fall 2023, MATH 583: High and Low Dimensional Statistics:
MWF 11:00-11:50 Neckers 0156. Webpage:
(http://parker.ad.siu.edu/Olive/M583hdim.html).

Text: Olive, D.J. (2023), *Prediction and Statistical Learning*, online course notes:
(http://parker.ad.siu.edu/Olive/slearnbk.pdf).

Suppose there are n individuals or objects, and p measurements $\boldsymbol{x}_i$ are made on the ith object. For regression, the response variable Y is of interest. For low dimensional statistics, $n \geq 10p$, say. For high dimensional statistics, $n < 5p$ say, and often p is much larger than n: $p >> n$. Emphasis will be on high dimensional regression methods such as multiple linear regression, binary regression, binomial regression, Poisson regression, Weibull regression, and Cox Proportional hazards regression. Classification and outlier detection will also be covered. Lasso, one component partial least squares, partial least squares, principal component regression, and marginal maximum likelihood estimators will be covered.

This course pairs well with Math 586: Statistical Learning. One method for high dimensional regression is to use variable selection or model selection, e.g. lasso variable selection or forward selection, to get a sparse fitted model: reduce the number of predictors from p to k where $n \geq 10k$. In other words, reduce the high dimensional problem to a low dimensional problem, and do inference using data splitting. A second method for inference also reduces a high dimensional problem to a low dimensional problem, for example for testing $H_0 : \beta_i = 0$, $i = 1, ..., p$ where $p >> n$ is possible. See the paper below.

A lot of the material is not yet published, so we will use some preprints such as Olive, D.J., and Zhang, L. (2023), "One Component Partial Least Squares, High Dimensional Regression, Data Splitting, and the Multitude of Models," at (http://parker.ad.siu.edu/Olive/ppopls.pdf). I expect at least three Ph.D. dissertations from this material.

The *prerequisites* for this class are Math 483 (calculus based statistics) and Math 221 (linear algebra: eigenvalues, eigenvectors, inverse matrices, et cetera). A course in regression would be useful (my Math 473, 485, or 485 cover regression models). *Heavy use of the computer package R* will be made. This course may be the fifth most difficult course that I teach (Math 581 and likely Math 580, 582, and 584 are harder.)

Other references.

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, New York, NY.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015), *Statistical Learning with Sparsity: the Lasso and Generalizations*, CRC Press Taylor & Francis, Boca Raton, FL.

Hu, J., and Bai, Z. (2015), "A Review of 20 Years of Naive Tests of Significance for High-Dimensional Mean Vectors and Covariance Matrices," *Science China Mathematics*, 55, online.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013, 2021), *An Introduction to Statistical Learning With Applications in R*, 1st and 2nd ed., Springer, New York, NY.

Koch, I. (2014), *Analysis of Multivariate and High-Dimensional Data*, Cambridge University Press, New York, NY.

Office: (J.W.) Neckers 261 (wing A) *Phone:* (618)-453-6566
email: dolive@siu.edu *Office hours: MWThF TBA*
I am also available by appointment and on a walkin basis, eg before class.

This course covers high dimensional statistics with emphasis on regression. Lasso, one component partial least squares, and marginal maximum likelihood estimators with data splitting will be important. Cumulative Final: TBA in week of Dec. TBA.

The grading and syllabus below are tentative. (Drop day in Friday, TBA, with advisor, Sunday, TBA online.)

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. One HW may be turned in one class period late with 25% penalty except for the last week of classes. One or more sheets of notes will be allowed on quizzes and exams. A calculator is permitted. I sometimes give $A-$, $B+$, $B-$, and $C+$.

HW	300	drop 1 HW	Quizzes	100	
exam1	100	exam 2	100	exam 3	100
final	300			total	1000
min. grade	points	min. grade	points	min. grade	points
A	900-1000	B	800-899	C	700-799
D	550-699				

Very Tentative

Week of	MON	WED	FRI
Aug 22	3.1	3.1	3.4
Aug 29	3.4	3.6, 3.12, HW1	3.12, Q1
Sept 5	no class	3.12,HW2	3.13, Q2
Sept 12	3.13	3.6, HW3	3.6, Q3
Sept 19	3.7	3.9, HW4	3.10, Q4
Sept 26	2.3	Exam 1	2.4
Oct 3	2.5	1.4, HW5	4.1, Q5
Oct 10	4.2	4.3, HW6	4.3, Q6
Oct 17	4.4	4.4, HW7	4.6, Q7
Oct 24	4.6	4.9, HW8	4.9, Q8
Oct 31	4.10	Exam 2	4.11
Nov 7	4.11	5.1, HW9,Q9	5.2
Nov 14	5.2	5.3,HW10	5.3, Q10
Nov 21	no class	no class	no class
Nov 28	6.1	6.1, HW11	6.2, Q11
Dec 5	6.2	Exam 3	review