

Successful Data Mining in Practice

A one-day seminar taught by

Richard De Veaux Ph.D., ASA Fellow

Sponsored by

**The Cleveland Chapter
of the
American Statistical Association**



Thirty-second seminar in a series on
Tools for Regaining the Competitive Edge

dedicated to the memory of Jerry Senturia, ASA Fellow, founder of this seminar series

Holiday Inn, 6001 Rockside Road (I-77, exit 155), Independence, OH

Friday, April 13, 2018

About the Instructor



Richard De Veaux, Ph.D. (Dick) is C. Carlisle and Margaret Tippit Professor of Statistics at Williams College. He holds degrees in Civil Engineering (B.S.E. Princeton), Mathematics (A.B. Princeton), Dance Education (M.A. Stanford) and Statistics (Ph.D., Stanford), where he studied with Persi Diaconis.

Dick has taught at the Wharton School and the Engineering School at Princeton and has been a visiting researcher at INRA in Montpellier and a visiting professor at the University of Paris Descartes (Paris V). De Veaux has won numerous teaching awards from the Engineering Council at Princeton.

He has won both the Wilcoxon and Shewell (twice) awards from the American Society for Quality, is a fellow of the American Statistical Association (ASA) and an elected member of the ISI. In 2006-2007 he was the William R. Kenan Jr. Visiting Professor for Distinguished Teaching at Princeton. In 2008 he was named the Statistician of the Year by the Boston Chapter of the ASA. He has served on the Board of Directors of the ASA and is current chair of the Section on Statistical Learning and Data Science.

Dick has been a consultant for nearly 30 years for such Fortune 500 companies as Hewlett-Packard, Alcoa, Dupont, American Express, Bank One, GlaxoSmithKline, Pillsbury, SanofiPasteur and General Electric. He holds two U.S. patents and is the author of more than 40 refereed journal articles.

He once helped Mickey Hart of the Grateful Dead with the question "How many drummers are there in the world" and for that he is known as the "Official Statistician of the Grateful Dead".

(continued in next column)

About the Seminar

Successful Data Mining in Practice

This one-day seminar is a practical introduction to and an overview of the techniques and strategies of data mining. While I will discuss the models in detail, the course will be application oriented rather than theoretical.

Many of the standard techniques of data mining, including a review of modern model selection strategies for multiple regression such as the lasso, elastic net etc will be presented. In addition we'll cover classification and regression trees, neural networks, principal component regression, Naïve Bayes, bagging, and boosting.

The course will be problem solving based, using real case studies from industry to illustrate which methods work well, when and why. We will emphasize problem formulation, the challenges of the data, and the communication back to decision makers in order to effect maximum impact in the organization.

No prerequisites other than a knowledge of the basics of regression are assumed. The applications will come from a wide variety of industries and include some applications from my personal experiences as a consultant for companies that deal with such topics as financial services, chemical processing, pharmaceuticals, and insurance.

About the Instructor (cont.)

Dick is the co-author, with Paul Velleman and David Bock, of the critically acclaimed textbooks "*Intro Stats*", "*Stats: Modeling the World*" and "*Stats: Data and Models*" and with Norean Sharpe and Paul Velleman of "*Business Statistics*", and "*Business Statistics: A First Course*", all published by Pearson.

His hobbies include cycling, swimming, singing -- and dancing (he was once a professional dancer and taught Modern Dance during Winter Study at Williams). He also teaches a course called "The history, geography and economics of the wines of France". He is the father of four: two boys and two girls ages 28, 26, 24 and 22.

Reminder

Full participation in this seminar can result in Recertification Units (R.U.s) earned towards maintenance of ASQ certification requirements.

Friday, April 13, 2018

Agenda

- 8:30-9:00 Registration and continental breakfast
- 9:00-10:20 OVERVIEW OF DATA MINING
Why do data mining?
Application areas
Conceptual challenges
Types of models: *predictive* (classification, regression, time series); *descriptive* (clustering, association detection, sequence detection)
Data mining myths
Practical challenges2
- 10:20-10:40 Refreshment Break
- 10:40-12:00 GETTING STARTED
Graphics
Stating the business problem
Beginning models – regression variants
Penalized regression
Logistic regression
Testing and training – Cross-validation
- 12:00-1:00 Lunch Break
- 1:00-2:30 MODELS
Trees
Smoothing
Additive models
Projection Pursuit – Neural networks
Deep learning
Collaborative filtering
Naïve Bayes
- 2:30-2:50 Refreshment Break
- 2:50-4:30 COMBINING MODELS/SUMMARY
Bagging
Boosting
Model comparisons
Clustering/Unsupervised learning
- WHAT CAN GO WRONG
Overfitting
Model limitations
- SUMMARY
Case Studies
Lessons learned
Where to go from here?

Lunch Selections

Please indicate choice on Registration Form

- ✦ Parmesan Encrusted Chicken
- ✦ Cheese Ravioli with Vodka Blush Sauce

Directions:

The Holiday Inn is located at 6001 Rockside Road, Independence, OH 44131

from the AIRPORT:

Take Interstate 480 East to I-77 South, exit at Independence-Rockside Road

from MAJOR HIGHWAYS: Traveling on the Ohio Turnpike (I-80), exit the Turnpike at Exit 11. Head north on Ohio Route 21 to Rockside Road (approx. 8 miles). Proceed west on Rockside Road to the I-77 interchange. The hotel is on the right hand side of the road.

from the Northwest (Toledo Area): Travel east on I-90 to the I-77 interchange. Head south on I-77 to the Rockside Road Exit and go east (left) on Rockside Road to the hotel.

from the Northeast (Erie, PA/NY Thruway): Travel west on I-90 to the I-271 interchange. Proceed south on I-271 to the I-480 west interchange and head west on I-480 to the I-77 south interchange. Then travel south on I-77 to the Rockside Road Exit and go east on Rockside Road to the hotel.

from the Southwest (Columbus Area): Travel north on I-71 to the I-480 interchange. Then head east on I-480 to I-77 south Exit. Take I-77 south to the Rockside Road Exit (1st exit) and turn left on Rockside Road to the hotel.

from the South (Akron/Canton Area): Travel north on I-77 to the Rockside Road Exit, turn left to hotel.

from the South/Southeast (Youngstown/Pittsburgh Area):

Travel west on the Ohio Turnpike (I-80) to Exit 13 (I-480). Head west on I-480 to the I-77 South interchange. Then travel south on I-77 to the Rockside Road Exit. Go east on Rockside Road to the hotel.

***Alternate Route:** Travel west on the Ohio Turnpike to Exit 11 (Ohio Route 21). Then head north on Route 21 to the Rockside Road (approx. 8 miles). Proceed west on Rockside Road to the I-77 interchange. The hotel is on the right hand side of the road.

Note: Traffic at I-77, exit 155 is very heavy between 7:45 and 8:30 AM.

Accommodations

Holiday Inn

6001 Rockside Road
Independence, OH 44131
(216) 524-8050 x298

There is a block of rooms reserved for \$95.00 + tax, double or single occupancy. Reduced rate is available until March 25. Be sure to mention you are with the Cleveland Chapter of ASA.

Complementary shuttle service is available from the Cleveland Hopkins Airport from 6:00 AM to 10:30 PM. Phones are found in the Baggage Claim area, lower level.

Register by credit card [online](#)

Or to pay by check, use the **Registration Form below**

No new reservations at the door

Make check payable to: Cleveland Chapter, ASA – Mail check and this form (one per registrant, please) to:

Jerry Moreno
JCU Dept of MT/CS
1 John Carroll Blvd.
University Heights, OH 44118
(216) 397-4681 email: moreno@jcu.edu

Deadline
April 2, 2018
Refunds cannot be
made after that date

Name (please print): _____ Day phone: () _____

Position: _____ Company: _____

Mailing address: _____ City: _____

State: _____ Zip: _____ Email address: _____

Until 3/19

3/20-4/2

| | | | |
|-----------------------|--|--------------------------------|--|
| Full Time Student | \$50 <input type="checkbox"/> | \$80 <input type="checkbox"/> | Faculty Signature: _____ |
| ASA/ASQ | \$170 <input type="checkbox"/> | \$220 <input type="checkbox"/> | Please indicate membership |
| Non-member | \$200 <input type="checkbox"/> | \$250 <input type="checkbox"/> | <input type="checkbox"/> ASA Chapter: _____ |
| Indicate lunch choice | <input type="checkbox"/> Parmesan Encrusted Chicken | | <input type="checkbox"/> ASQ Division: _____ |
| | <input type="checkbox"/> Cheese Ravioli with Vodka Blush Sauce | | |