Advanced Statistics for Business Analytics (BAX 442)

Instructor
Prof. Prasad A. Naik
LinkedIn: https://www.linkedin.com/in/prasad-a-naik-a868a459/. (Please feel free to connect.)
Emails: panaik007@gmail.com
Mobile: (530) 304 1348

Brief Bio
I graduated as a Chemical Engineer from the Institute of Chemical Technology (ICT) and then worked at Dorr-Oliver Ltd. in B2B sales. Then, I studied MBA at the Indian Institute of Management Calcutta (IIMC) and worked at GlaxoSmithKline (now Unilever) in sales management to manage wholesalers, retailers, and salesforce and later in brand management to launch new products and build strong brands. I earned a PhD in Marketing from the University of Florida, and I am serving as a faculty member at the UC Davis for over 25 years.

As a Professor, I build knowledge and talent. My research advances knowledge. Teaching develops talented students like you. Research and teaching are two sides of the same coin. Both activities change beliefs and complement each other.

Class Meetings
- Venue: Cotchett Law Center, Room LL01
- Dates:
  - January: 9, 16, 23, 30
  - February: 6, 13 (Midterm Exam), 20, 27
  - March: 6, 13, 20 (Final Exam)
- Timings:
  - 11:00 am through 3:00 pm (Section 1)
  - 4:00 pm through 8:00 pm (Section 2)
- Office Hours:
  - Via Zoom
  - Email me to arrange suitable days/times for zoom calls.

Audio Description:

Course Description
This course introduces cross-sectional and time series analysis to solve real-world business problems. To answer your questions 24 x 7, ask Athena. Click to use it.

Class 1 begins with the review of the linear regression model in a vector-matrix notation. In the first two hours, we will build our own R code from first principles and compare it with the pre-built lm() code in R. If they match exactly (i.e., not just approximately), you really know the linear regression model. If not, you know the usage of the pre-built code, but not the knowledge to build it yourself. The second layer of knowledge comes from learning how to build a "function" in R — that knowledge helps you code anything accurately and efficiently.

Armed with these learnings, in the last two hours of Class 1, we apply the linear model to the central marketing application of New Product Design. We shall study the three phases: descriptive, predictive, and prescriptive modeling. Based on the descriptive model, we estimate the willingness-to-pay (WTP) for the various features; the predictive model forecasts the preferences for various feature combinations; and, finally, the prescriptive model helps discover the optimal price to charge for a planned feature set of the new product.

HW1 presents you an opportunity to practice how to code a function for optimal pricing of new products that takes the data inputs and creates the desired outputs: WTP, optimal share, optimal price, and optimal profit. (The correct version of this function will also be used in HW2.)

Class 2 aims to understand the concept of confidence intervals. We will learn two methods —Bootstrap Regression and Monte Carlo Simulation— to find confidence intervals of *any* given function of the model's parameters. I will illustrate the application based on the smart phone data. In HW2, you will apply this knowledge to find the confidence intervals of WTP for the features of the new product you recommended in Class 1.

In the second half of Class 2, you will learn Principal Components Regression. Then we apply it to study another marketing application: visualizing competition. I will illustrate perceptual mapping based on the data on beer brands. In HW3, your team will practice this content using cars data to create the perceptual maps for the focal car brand (Infinity) and recommend the marketing strategy.

In linear regression, we need more observations than the number of variables. But what if the number of variables exceeds the sample size ($p > N$)? Linear regression fails in this case. Class 3 covers the three methods —Ridge regression, Lasso regression, and Elastic Net regression—that facilitate the estimation in the presence of more variables than the sample size. We will apply these methods to the Cars data and identify the Top 3 features of cars that influence consumers' preferences.

Additionally, the linear regression needs the model to be linear-in-parameters. Within this framework, how should we incorporate nonlinearity? For example, advertising increases sales, and more advertising further increases sales but less than proportionally. To incorporate nonlinear effects, we shall learn how to include polynomial functions, variable transformations, and variable interactions in linear-in-parameters models. HW4 presents you an opportunity to apply these ideas in the context of an important application in marketing called Marketing Mix Models (MMM). This MMM technology is becoming even more important (e.g., Google launched Meridian) because regulators enforce restrictions on tracking individual-level data.

Class 4 covers Nonlinear and Nonparametric Regressions. All the previous models assumed that the regression models are linear-in-parameters. But what if it is not? Moreover, what if models have no parameters at all? You will study the estimation, inference, prediction, and interpretation of nonlinear models with and without parameters. In HW5, you will predict nonlinear sales growth of products using Tri-Logistic Regression model (i.e., three S-shaped curves on top of each other unlike the single S-shaped logistic model).

In Class 5, we shall learn how to group the customers in multiple segments and fit a regression model within each group. To this end, we can group customers using K-Means cluster analysis and then fit regression models (sequentially). Alternatively, we can group customers and fit regression models simultaneously via the Mixture Regression model. In HW6, you will explore the central question, how many segments to retain?

In Class 6, the Midterm Exam will be based on the content in the above five classes. The exam will be in-class, multiple choice questions, with no access to books, notes, Internet, Smartphones, or Chat-GPT. The exam will be held from 11 am to 12 pm for the Section 1 and 7 pm to 8 pm for the Section 2. Note that there's no make-up exam – so please don't miss it.

After the exam, we will study the Seemingly Unrelated Regression, which extends the linear model to multiple dependent variables (e.g., online sales and offline sales in multichannel marketing applications). This topic completes the study of cross-sectional regression models in this course.

Complementing the analysis of cross-sectional data, we shall focus on Time Series Analysis for the remainder of the quarter. In time series analysis, the past dependent variable influences the present and the present impacts the future. This inter-temporal dependence is the main difference between cross-sectional versus time-series analysis. Given time-series variables (e.g., sales, inflation, weather), we shall learn how to separate the signal from noise, decompose this signal into level, trend, and seasonal components, and forecast future outcomes together with the associated uncertainty.

Class 7 covers time-series decomposition, Holt-Winters (HW) Filter, and ARIMA forecasting.
In the acronym ARIMA, "AR" refers to an auto-regressive process, which means the dependent variable depends on its own past values. "MA" refers to the moving average process, which means the dependent variable depends on the past values of its error terms. Thus, we extract information from past values of both the dependent variable and the error terms to improve forecasting. But both AR and MA processes require "stationarity," which means no trends are present in the time series. However, most time series exhibit increasing or decreasing trends. Hence, we remove these trends by differencing the dependent variable a few times. This differencing is called an integrated component and denoted by "I" in the ARIMA name. Like the decomposition models and HW Filter, the parameters of ARIMA models are constant over time. So forecasting via these approaches work in stable markets.

In Classes 8 and 9, we will learn how to forecast when the parameters themselves evolve over time (i.e., dynamic markets). To this end, we use Markov Switching Regression and State Space Models. The estimation approach is based on the Kalman Filter (KF), which NASA used to launch man on moon; the Department of Defense used it in anti-aircraft gunfire control. The KF also offers navigational guidance to spacecraft, aircrafts, ships, cars, and you. Although you may not know, you use it every day via the GPS app on your smartphones. I pioneered the application of KF in Marketing — click here to read it.

In Class 10, I will arrange an industry speaker to present a guest lecture. Your attendance is mandatory. If absent, 10 points will be deducted from your final score. After the guest lecture, we reserve time to catch up on the course material, review concepts, and discuss the advancing frontiers of statistics and business.

The Final Exam will be based on the content across all the above classes (i.e., comprehensive). The exam will be in-class, multiple choice questions, with no access to books, notes, Internet, Smartphones, or Chat-GPT. The exam timings will be announced later. As before, there's no make-up exam – so please don't miss it.

Grading

- Midterm Exam (25%)
- Homework (35%) – one least scoring HW will be dropped.
- Final Exam (40%)

Textbooks

1. **ISL**
   - *An Introduction to Statistical Learning: With Applications in R*, 7th Edition.
   - Download at https://trevorhastie.github.io/ISLR/ISLR%20Seventh%20Printing.pdf

2. **FPP**
   - *Forecasting: Principles and Practice* by Rob J. Hyndman and George Athanasopoulos.
   - Access online textbook at https://otexts.org/fpp2/
   - You can buy the printed version from Amazon: https://www.amazon.com/dp/0987507117?tag=otexts-20

3. **SSM**
   - *State Space Models*, Chapter 9, in the previous edition of *Forecasting: Principles and Practice*
   - Download the PDF from the course page on Canvas

Tentative Teaching Plan: If and when the syllabus changes, I will post the revised version on Canvas.

Class	Date	Topics	Readings	HWs
1	Jan 9	Course Overview Cross-sectional Analysis Begins (Classes 1 – 5) Linear Model in Vector-Matrix form • Estimation & Inference • Build LM code from first principles • Writing a "function" in R LM Application to Design New Products • Descriptive model (willing-to-pay) • Predictive model (preferences) • Prescriptive model (optimal price)	Class 1.pdf ISL Ch. 3	HW1: Build a function in R for WTP, shares, price, and profit Due: Jan 16th
2	Jan 16	Confidence Intervals for Anything • Bootstrap and Monte Carlo • Smartphone Application Principal Components Regression • Dimension Reduction • Perceptual Mapping Application (Beer Brands)	ISL Ch. 5.2 ISL Ch. 6.3.1 ISL Ch. 10.2	HW2: Find CIs for WTP Due: Jan 23rd HW3: Car Brands Map Due: Jan 30th

3	Jan 23	Penalized Regressions
		• Features exceed sample size (p > N)
		  ○ Ridge Regression
		  ○ Lasso Regression
		  ○ Elastic Net Regression
		• Car Brands Application
			ISL Ch. 6.2.1
			ISL Ch. 6.2.2	

		Nonlinearity via Linear Models • Polynomial, Transformation, Interaction • Multimedia Marketing Mix (MMM) Application	ISL Ch. 7.1 ISL Ch. 3.6.4 ISL Ch. 7.4.1	HW4: MMM Due: Feb 6th
4	Jan 30	Nonlinear Regression (Parametric) • NLS or Optim in R • Multi-logistic Sales Growth Application Nonparametric Regression • Loess • GAM	 ISL Ch. 7.6 ISL Ch. 7.7	HW5: Sales Growth Models Due: Feb 20th
5	Feb 6	Segmentation + Regression Part I: Sequential • K-means Clustering, then Regression • Smartphone Customer-level Application Part II: Simultaneous • Mixture Regression Models ○ Classification and Regression • Smartphone Customer-level Application Catch up, Review, Q&As for Midterm Exam • Come prepared with your questions	 ISL Ch. 10.5	HW6: How many segments to retain? Due: Feb 27th

6	Feb 13	Midterm Exam
		• No make-up exam (don't miss it)
		• Content on classes 1 through 5
		• In-class, multiple choice
		• No books, notes, Internet, Chat-GPT
		• Timings
		  ○ Section 1: 11 am to 12 pm
		  ○ Section 2: 7 pm to 8 pm
		
		Class 6 continues after the midterm exam
		
		Seemingly Unrelated Regression (SUR)
		• Multichannel Marketing Application
		
		Cross-sectional Analysis Ends and Time Series Analysis Begins (Classes 6 – 10)
		
		Overview
		• Forecasting Process
		• Correlation and Causality
		• Spurious Regression
			FPP Ch. 1
			FPP Ch. 5.3
			FPP Ch. 5.9
				No HWs due this week

		• Qualitative Forecasting • Quantitative Forecasting	FPP Ch. 4.3 FPP Ch. 5.6	
7	Feb 20	Decomposition Models • Levels, Trends, Seasonality • Holt-Winters Filter ARIMA Forecasting • AR process • MA process • Differencing • ACF and PACF Breakout Work (in-class)	FPP Ch. 6 FPP Ch. 8	HW7: Quantify Growth and Seasonality Due: Mar 6th HW8 ARIMA Forecasts Due: Mar 13th
8	Feb 27	Markov Switching Regressions • Identify Multiple Marketing Regimes State Space Models • Unification of models on Levels, Trends, Seasonality, and ARIMA	 SSM Ch. 9.1 SSM Ch. 9.2 SSM Ch. 9.4	No HW No HW
9	Mar 6	State Space Models • Kalman Filter (KF) • Time-varying coefficients in LM State Space Models • Maximum Likelihood Estimation (MLE) • Estimate LM with time-varying coefficients • Dynamic MMM Application Breakout Work (in-class) Time Series Analysis Ends Here	 SSM Ch. 9.3 SSM Ch. 9.6 SSM Ch. 9.3 on p. 99	No HW
10	Mar 13	Guest Lecture (Attendance Mandatory) Catch up, Review, Q&As for the Final Exam • Come prepared with your questions. • Ask anything you always wanted to know in statistics, business, and life.		No HW
11	Mar 20	Final Exam		

Course Conduct

1. Attendance in the first and the last class (Jan 9th and Mar 13) is mandatory without exceptions. If absent, you lose 10 points per class from your final score.

2. No make-up for the Midterm Exam (Feb 13th) and the Final Exam (Mar 20th). Mark your calendars.

3. Learning demands full attention and concentration. Please do not check emails, surf websites, use smartphones, or send texts during the class lectures. These distractions hinder your learning.

4. Absences. We meet only 10 times over the quarter and so each class contributes 10%. So avoid absences. But, if you must miss a class (other than Jan 9th and Mar 13th), please let me know in advance via email (panaik007@gmail.com). Maximum absences of two classes this quarter are allowed. For additional absences, you lose 10 points per absent class from your final score.

5. Homework. HWs will be done in teams. Please form a team of 5 students before the end of Class 1. The tentative teaching plan indicates when the HWs are assigned and due. Submit your HWs on the due dates before the class begins.

University Policies

1. **Statement on Accommodation.** To seek accommodation for learning disabilities, visit the Student Disability Center (at https://sdc.ucdavis.edu/) and contact them at sdc@ucdavis.edu or 530-752-3184. Once you receive the Letter of Accommodation, submit it to me as soon as possible within the first two weeks of the course.

2. **Rights and Responsibilities.** Instructor and all students are expected to follow the UC Davis Principles of Community, which includes freedom of expression, rejection of discrimination, and other issues (for details, see https://diversity.ucdavis.edu/principles-community).

3. **Code of Academic Conduct.** You are required to uphold the University's Regulation 537 on Exams, Plagiarism, Unauthorized collaboration, Lying, Disruption, and other issues. Read the academic code of conduct at this link: http://sja.ucdavis.edu/files/cac.pdf. If you are in violation, you will be referred to the Office of Student Judicial Affairs.

4. **Safety and Emergency Preparedness.** Please familiarize with the campus Emergency Information at https://www.ucdavis.edu/emergency/. To get timely information and instructions about emergencies and situations on campus that may affect your safety, you may register at UC Davis Warn Me and Aggie Alert. In case of emergency in the classroom or in non-Davis locations, follow the instructions of your instructor.

5. **Special Statement on COVID-19.** As your teacher, I understand that you may face obstacles in meeting your academic goals. Your safety, health, and well-being are important to me. Please be aware that UC Davis provides a wide range of remote services such as counseling, tutoring, academic advising, and community building and engagement. If you feel your class performance is affected, please do not hesitate to contact me. I am committed to helping you to meet your learning objectives in this course.

***** End of Syllabus *****