

ROBERT COYIUTO PROFESSORIAL CHAIR IN STATISTICS

Survival Analysis in the World of Big Data

by John Eustaquio

With the influx of big data mostly in customer-oriented companies, data mining techniques became a common tool for gathering insights and derive patterns that are beneficial to reaching a company's objectives. A lot of analysis coming from big data deals with predicting time-to-event features like when is a customer is going to churn or the amount of time that a person will click the ad placed on social networking sites. The analysis of these time-to-event features is frequently dealt by Survival analysis. Although classical methods in survival analysis is frequently used for time-to-event data, it encounters some problems with big data due to high dimensionality and massive sample size. High dimensionality in survival models is a common problem with genomic data were the number of biomarkers is immensely higher than the number of subjects who experienced a certain disease. In customer lifetime modeling for credit card companies and subscription-based companies, the number of observations and number of time periods often far exceeds millions. In these scenarios, method for high dimensional and massive sample size data needs to be incorporated. The objective of this presentation is to discuss the current developments in dealing with high dimensional and massive sample size data in survival models.

ANGELES BUENAVENTURA PROFESSORIAL CHAIR IN STATISTICS

Scoring System for Graduate School Admissions using Logit Models

by Josefina Almeda

University administrators encounter difficulties in determining applicants who will achieve success in the graduate studies. Many are admitted but very few graduate. To help screen applicants, this study determined the predictors of student success and developed a performance score differentiating high-risk and low-risk applicants. The participants were College of Education students who completed and did not complete the Master of Arts in Education for the academic years 2005-2006 to 2011-2012. This excluded continuing students who have not reached the maximum residence of their degree programs. Logit models for master's students were estimated relating their personal, academic, and professional characteristics to their probability of completing the graduate program. Predictors that increased the probability of master's completion success were Master's Admission Test in Education (MATE) score, being a full-time student, with college academic honors, work as manager or supervisor, and works with a non-government organization. There was a slight net negative effect of the gap in years at the mean from college to master's program admission. Being a teacher has a lower probability of completing the graduate program than other types of occupation. The 0.1 probability cutoff value was the basis for determining if master's applicants will complete or will not complete the graduate program. The estimated master's logit models correctly predict the actual completion of the graduate program status 60.16% of the time. The estimated models serve to supplement the University administrators' admissions process.