SOCIETY OF STATISTICS, COMPUTER AND APPLICATIONS

(Celebrating 25 Years of its foundation)

Web - Workshop

"Linear Algebra, Matrix Theory and Linear Estimation"

Friday, 02 September 2022

Organizer

Simo Puntanen

Visiting Researcher, Tampere University, Finland

Time: 04:00 PM – 08:30 PM (India Time); 06:30 AM – 11:00 AM (EST, Canada);

05:30 AM – 10:00 AM (CST); 03:30 AM – 08:00 AM (PST);

Central European Time: 11:30 AM - 04:00 PM

REGISTRATION LINK

To register for the webinar, please click

https://us02web.zoom.us/meeting/register/tZErc-yopj4oGda-uwzzpoye9rIO8lT_m0DG

ABOUT THE WEB - WORKSHOP

Every thorough and disciplined course in Linear Estimation [Theory and Methods] – even at the beginner's level - rests on a good understanding about Linear Algebra. Linear algebra, very soon and quite naturally, ends into matrix manipulations of varying degrees! One wonders how these broad topics are so intricately interrelated.

Dear Participants, it is really a wonderful world of matrix manipulations to start with! You will get a glimpse of such tricks that come handy and, naturally, are very useful.

Coming to think of Linear Regression [under usual assumptions], we tend to believe that we are at the top of the world ! We teach this topic so comfortably and easily -- again and again. Wait ! Here is a smart student asking a critical question indeed ! AND the whole thinking process gets radically changed !

Even a teacher wonders how valid and relevant is the quote:

"There is more to see (on the ground) than what we see (from the sky)." **We will show case this part of the many wonders!**

Again, at an intermediate level, students are introduced to linear random coefficient models and naturally variance components are brought in. However, with missing covariate values, the underlying estimation and inference aspects are worth studying.
Yes……one speaker will dwell on this topic.

Oh, Dear ! The Sum of Squares [SS] and its computation while working on the ANOVA Tables are so easy-going exercises and we all love to carry out the computations !

Stop and Behold ! It may NOT be that fun any more if you are dealing with some complicated data even with two factors in a simple model.
Yes….one of us will draw your attention to this aspect as well.

On a serious note, longitudinal data analysts have a lot to say ! They are engaged in meaningful understanding of the underlying models. And we have a specialist in this topic of research. He will appraise us about the ways to improve upon the existing models with adequate justifications and illustrative example(s).

After participating in this workshop, the young researchers would get an insight into the power of matrix theory, linear algebra and their usefulness in linear estimation. Curiosity about application of linear algebra concepts to real datasets would agitate the young minds. An understanding about the use of linear algebra in computers to solve real-world applications in machine learning, data science, statistics, and signal processing would develop.

Some Useful References

- Christensen, R. (2011). Plane Answers to Complex Questions: The Theory of Linear Models. 4th Edition. Springer, New York. DOI - http://dx.doi.org/10.1007/978-1-4419-9816-3
- Harville, D. A. (1997). Matrix Algebra from a Statistician's Perspective. Springer, New York. DOI - http://dx.doi.org/10.1007/b98818
- Rao, C. R. and Mitra, S. K. (1971). Generalized Inverse of Matrices and its Applications. Wiley, New York.
- Rao, C. R. (1973). Linear Statistical Inference and its Applications. John Wiley & Sons.
- Searle, S. R. (1971). Linear Models. John Wiley & Sons.

SCHEDULE

Time	Topic	Speaker
4:00 PM - 4:10 PM	Welcome Address	Vinod K Gupta
4:10 PM - 4:25 PM	About the Workshop	Simo Puntanen
4:25 PM - 5:10 PM	MATRIX TRICKS YOU CAN'T LIVE WITHOUT IN THE WORLD OF LINEAR STATISTICAL MODELS	Simo Puntanen*
5:10 PM - 5:55 PM	UNDERSTANDING LINEAR AND QUADRATIC REGRESSION IN THE LIGHT OF DE LA GARZA PHENOMENON	Bikas K Sinha*
5:55 PM - 6:40 PM	ANOVA SSs AND PROPORTIONAL SUBCLASS NUMBERS	Lynn Roy LaMotte*
6:40 PM - 7:25 PM	ON THE IMPROVED ESTIMATION OF NORMAL MIXTURE COMPONENTS FOR LONGITUDINAL DATA	Tapio Nummi*
7:25 PM - 8:10 PM	INFERENCE IN SIMPLE LINEAR RANDOM COEFFICIENT MODEL WITH POSSIBLY MISSING COVARIATES	Julia Volaufova*
8:10 PM - 8:25 PM	Summary of the Workshop	Simo Puntanen
8:25 PM to 8:30 PM	Vote of Thanks	A. Dhandapani

**** Note: 40 minutes for presentation and 5 minutes for discussion***

ABOUT THE SPEAKERS AND TOPICS

Simo Puntanen, is a retired Adjunct Professor at Tampere University, Finland. He did his Ph.D. at Tampere University, in 1987, where working up till retirement in 2013. Research interests include matrix methods in statistics, with particular emphasis on linear statistical models. Organizer/co-organizer of numerous conferences and workshops, among them a co-founding member of the series of the International Workshop on Matrices and Statistics, IWMS. Tampere has been his headquarter but he's never been unwilling to accept research visits to India, Canada, New Zealand and Poland, for example.
Link to publications: https://homepages.tuni.fi/simo.puntanen/papers-sjp.html

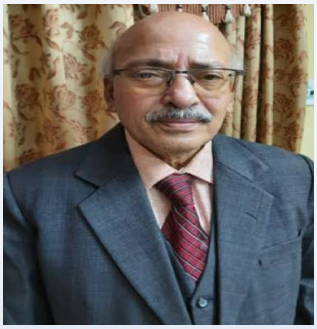
Prof. Puntanen will be talking on…

MATRIX TRICKS YOU CAN'T LIVE WITHOUT IN THE WORLD OF LINEAR STATISTICAL MODELS

Matrix tricks that we trust in - that's the motto of this talk. In other words, we go through some matrix tricks that we have found particularly helpful when dealing with linear statistical models. Some of them appear in the literature in a slightly hidden form and we try to upgrade them into a business class, so to say. In teaching linear statistical models to first-year graduate students or to final-year undergraduate students there is no way to proceed smoothly without matrices and related concepts of linear algebra; their use is really essential. Our experience is that making some particular matrix tricks very familiar to students can substantially increase their insight into linear statistical models (as well as into multivariate statistical analysis). In matrix algebra, there are handy, sometimes even very simple "tricks" which simplify and clarify the treatment of a problem - both for the student and for the Professor. Puntanen, Styan, and Isotalo, knowing just what is expected of authors, would like to agree with P. G. Wodehouse and apologize for childhoods that were as normal as rice-pudding and lives that consisted of little more than sitting in front of the laptop and cursing a bit.

References

- Puntanen, Simo, Styan, George P. H. and Isotalo, Jarkko (2011). Matrix Tricks for Linear Statistical Models, Our Personal Top Twenty. Springer, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-10473-2
- Puntanen, Simo, Styan, George P. H. and Isotalo, Jarkko (2013). Formulas Useful for Linear Regression Analysis and Related Matrix Theory, It's Only Formulas But We Like Them. Springer, Heidelberg. DOI: https://dx.doi.org/10.1007/978-3-642-32931-9



Prof. Bikas K. Sinha, Retired Professor, ISI, Kolkata, India. He did his Ph.D. (1972) in Statistics from Calcutta University. Prof. Sinha has established himself as a statistician of international repute, being attached primarily with the Indian Statistical Institute, Kolkata. With more than 110 research collaborators worldwide and more than 160 peer-reviewed journal publications, Prof. Sinha is versatile in both statistical theory and applications in different areas. He has participated in many workshops and conferences on behalf of SSCA and other academic organizations within India and elsewhere.

Prof. Sinha will be talking on...

UNDERSTANDING LINEAR AND QUADRATIC REGRESSION IN THE LIGHT OF DE LA GARZA PHENOMENON

The celebrated de La Garza Phenomenon [AMS, 1954] deals with continuous or approximate designs and is fundamental to the study of optimal polynomial regression designs. However, in exact small sample cases, its validity may be in question. We illustrate this phenomenon with appropriate theory and examples. This is an instructional talk for teachers and young researchers interested in the area of regression designs.

Lynn Roy Lamotte, is Professor Emeritus in the Biostatistics Program at the LSU Health - New Orleans School of Public Health. His earned degrees are all from Texas A & M University, culminating in a Ph.D. in statistics (1969). He was elected Fellow of the American Statistical Association in 1985. His research efforts have focused mainly on linear statistical models, broadly, ranging from computational methods to basic results on admissibility among linear estimators, including methodology for random effects and variance components. Recent work, jointly with Jeff Wells, a forensic entomologist, and others, has developed methods for inverse prediction in mixed multivariate models for inference on time since death based on insect evidence.

Prof. Lynn will be talking on...

ANOVA SSs AND PROPORTIONAL SUBCLASS NUMBERS

In balanced models that include only main effects and interaction effects of factors A and B, the A sum of squares, SSA (unadjusted for B, or Type I SS) tests A main effects. The same is true of SSIIA, for A main effects adjusted for B main effects (Type II SS): in fact SSA = SSIIA. Since the beginnings of ANOVA, this fact has inspired hope that somehow such a simple, easily-computed expression could be found for a SS to test A main effects in unbalanced models. It has been widely believed and taught that the nice properties of SSA held also in unbalanced models with proportional subclass numbers (psn). Snedecor and Cox (1935) suggested, therefore, that subclass numbers n_{ij} be replaced by $n_{i.}\ n_{.j}/n..$ and that then SSA should be computed as a single-factor sum of squares. When SAS introduced Type III SSs, Goodnight (1976) validated the method by asserting that "[w]hen no missing cells exist in a factorial model, Type III SS will concide with Yate's [sic] [(Yates 1934)] weighted squares of means technique." SAS's description of the Type III method was (is) difficult to translate into its properties as a test statistic. Users of statistics in several fields were suspicious of it, and over the last few decades there has grown up a movement that asserts that Type II SS should be used instead. The purpose of this talk is to show exactly what SSA and SSIIA test in several different unbalanced settings, including psn. You will see that neither tests the right set of linear contrasts, even with psn, and that, except under very special conditions, they test no A main effects contrasts at all.

References

- Goodnight, J. H. (1976). The General Linear Models Procedure. Proceedings of the First International SAS User's Group. SAS Institute, Cary, NC.
- Snedecor, G. W. and Cochran, G. M. (1935). Disproportionate subclass numbers in tables of multiple classification. Research Bulletin No. 180, Agricultural Experiment Station, Iowa State College, Ames, Iowa.
- Yates, F. (1934). The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes. Journal of the American Statistical Association, 29(185), 51-66.

Tapio Nummi, is Adjunct Professor, in the Faculty of Information Technology and Communication Sciences at the Tampere University and has been a principal investigator in several Academy of Finland research projects on statistics, forestry and health. His field of application in statistics includes forestry, health, medicine and social sciences. His research interest in statistics includes methods for the analysis of longitudinal data, growth curves, mixed models, nonparametric regression, mixture regression and trajectory analysis. He has contributed 160 scientific publications and about half of them in refereed journals and books. Prof. Tapio will be talking on...

ON THE IMPROVED ESTIMATION OF NORMAL MIXTURE COMPONENTS FOR LONGITUDINAL DATA

In many modeling applications a finite normal mixture (see e.g. Everitt and Hand, 1981) as well as its extensions, such as mixture skew-normal distribution (Lin et al. 2007b) or mixture t-distribution (Lin et al. 2007a), provides a sensible model for the data at hand. However, finding the best possible component distribution among many competing alternatives can be a very complicated and challenging task. For example, the number of components needed may depend on the sample size, assumed component distribution as well as the possible transformation (scale of measurements) applied. Actually, two slightly different goals can be identified: Approximation of the distribution of the response variable and identification of the number of sub-populations. The focus of this talk is on so-called trajectory analysis (TA) that is an application of finite mixture modeling for longitudinal data (Nagin, 1999 and 2005). We propose a method that is based on the scaled Box-Cox transformation (Spitzer, 1984 and Gurka et al. 2006) that makes the likelihood based inference possible over transformed responses and the actual analyses are then based on the components of transformed normal mixtures. The proposed approach has several advantages. First, the theory of TA with normal mixtures is well established and many implementations as software packages already exist. Second, a suitable transformation can reduce the risk of generating the excess number of mixture groups. The data analytic part of the talk is based on real data applications of birth weight, trajectories of alcohol consumption as well as simulation experiments.

References

- Everitt, B. and Hand, D. (1981). Finite Mixture Distributions. London: Chapman and Hall.
- Gurka, M., Edwards, L., Muller, K. and Kupper, L. (2006). Extending the Box-Cox Transformation to the Linear Mixed Model. Journal of the Royal Statistical Society, A (Statistics in Society), 169(2), 273-288.
- Lin, T., Lee, J. and Hsieh, W. J. (2007a). Robust Mixture Modeling using the Skew t-Distribution. Statistics and Computing, 17, 81-92.
- Lin, T., Lee, J. and Yen, S. (2007b). Finite Mixture Modeling using the Skew Normal Distribution. Statistica Sinica, 17, 909-927.

Julia Volaufova, was born, raised, and educated in Slovakia, part of former Czechoslovakia. After a career in biostatistics, she retired in September 2017 when she became Professor Emerita at the Biostatistics Program of School of Public Health, LSU - Health New Orleans. Her primary interest in methodological research includes mixed linear models, regression analysis, and statistical hypothesis testing. She was also highly

active in applications of statistical methodology in the area of food intake, dietary supplements, calorie restriction, body weight, and diabetes. She is a fellow of ASA and lives partly in Portland, Oregon in the US and Bratislava, Slovakia.
(See also *https://en.wikipedia.org/wiki/Julia_Volaufova)*

Prof. Julia will be talking on…

INFERENCE IN SIMPLE LINEAR RANDOM COEFFICIENT MODEL WITH POSSIBLY MISSING COVARIATES

This talk is a brief overview, meant for broad mixed audience of students, young researchers, as well as statisticians very familiar with the presented topic. We shall deal here with a simple linear random coefficients (LRC) model. We assume that some of the observations of the covariates may be missing. Since approximately the mid 70s of the last century, estimation and prediction have been extensively studied in a variety of rather complex statistical models under the assumption that some observations are not available, they are missing. However, much less is available about statistical inference (hypotheses tests) when some of the covariates are missing. Here we extend and apply the distributional model for the missing covariates that depend on the completely observed covariates' values, which was introduced by Chen et. al (2008) in a linear fixed effects model setting. The primary interest here is to investigate properties of hypothesis tests about the mean parameters, constructed using maximum likelihood estimates obtained by the EM algorithm. We investigate the properties of tests that are commonly used in the classic LRC models. These tests include the Wald test, Score test, and the likelihood ratio test. When constructing the Wald and Score test statistics, we consider using both the Fisher information and its observed version to estimate the covariance matrix of the estimator of the hypothesized parameter. The accuracy of p-values and adjusted power of tests were studied using a simple simulation study, in which data were generated from a LRC model where among the covariates, there is only one for which some of the values for some sampling units are missing.

References

- Chen, Q., Ibrahim, J. G., Chen, M. H. and Senchaudhuri, P. (2008). Theory and Inference for Regression Models with Missing Responses and Covariates. Journal of Multivariate Analysis, 99(6), 1302-1331. https://doi.org/10.1016/j.jmva.2007.08.009

NOTE: Young researchers/Faculty Members desirous of getting a ***participation certificate*** may pay registration fee of **INR 100.00 (*Rupees One Hundred Only*).** Payments can be made through digital payments (On-line banking/ UPI/Google Pay etc.) to the following account :

Account Name	: **23rd Annual Conference of SSCA**
Account number	: **129510100135668**
Type of Account	: Savings
Bank Name	: Union Bank of India
Branch	: Attapur
IFSC Code	: **UBIN0812951**

V. K. Gupta
President

A. Dhandapani
Vice President

Society of Statistics, Computers and Applications (SSCA)

Website: ***https://ssca.org.in/***