

<p style="text-align: center;">NCCC-170 / USSES Joint Annual Meeting June 22 – 23, 2017 Cargill Innovation Campus, Elk River, MN</p>

Location: Cargill Innovation Campus, 10383 165th Ave NW, Elk River, MN 55330

<p style="text-align: center;">Technical Program</p>

Thursday, June 22, 2017

8:00 – 8:30 AM

Registration and Check-In

8:30 – 8:40 AM

Welcome and Introduction:

Dr. David Cook, PhD. Director of R&D and Managing Director of the Cargill Experiment Station.

8:40 – 9:20 AM

“The All Configurations Maximum Interaction F-test for Non-additivity and the hiddenf Package in R”

Jason Osborne, Dept. of Statistics, North Carolina State University

Chris Franck, Dept. of Statistics, Virginia Polytech Institute and State University

Dennis Boos, Dept. of Statistics, North Carolina State University

Bong Choi, SAS Institute

Abstract: A test is presented for a particular type of non-additivity in two-factor layouts. The term “hidden additivity” is introduced to describe cases where levels of one factor (B) can be placed into a smaller number of groups such that the effects of the other factor (A) vary across groups but are additive with factor B effects within groups. Tests of interaction between factors are developed using this notion of hidden additivity as an alternative hypothesis. Additionally, interval estimation for group-specific contrasts are developed using sets of most probable configurations of levels of factor B into groups. A software package in R (“hidden”, available on CRAN) provides tools for data analysis.

9:20 – 10:00 AM

“Possible Advantages of Matching Concepts in Agricultural Experiments with Blocks”

Elaine Sotherden and William Bridges, Dept. of Mathematical Sciences, Clemson University

Abstract: In many agricultural studies the experimental units are grouped into blocks based on covariates available when the experiment is planned. The treatments are randomly assigned to experimental units within the blocks, and the blocks are included in the model used to analyze the study. In some studies, additional covariates are measured while the experiment is ongoing. A question then becomes, can the information from the additional covariates be used for “post blocking” such that the treatment effect is more effectively measured (similar to matching in observational studies)? Issues that could impact this answer include: 1) how unbalanced are the new covariates among the treatments; 2) how much do the new covariates impact the response and what is the nature of the covariate impact on the response; 3) how much do the treatments influence the new covariates; 4) how much information is used or lost if the new covariates are used to improve the blocks; 5) exactly how is the “post blocking” done (some sort of matching, or some sort of ANCOVA); and so forth.

To obtain a very preliminary answer to this question, a study of sheep grazing endophyte infected fescue was considered. There were four treatments in the study and eight blocks of sheep (based on covariates related to body condition score). The sheep were on the treatments for approximately 140 days, and some additional covariates were measured during this period. Using these covariates to “post block” did in fact modify the estimated treatment effect, but we are still investigating properties of the new estimated treatment effect. Some simulation studies to address the question are underway and will be discussed.

10:00 – 10:20 AM

Morning Break

10:20 – 11:00 AM

“Pooling of variances: the skeleton in the mixed model closet?”

Philip Dixon, Department of Statistics, Iowa State University

Abstract: This talk is intended to solicit opinions and advice on an issue that I worry about intermittently: when is it appropriate to pool variances and when should that be avoided. The issue is not new: one instance is inference about the difference of means from two groups. If you pool, you get the two-sample t-test; if you do not, you have the Behrens-Fisher problem, which has various approximate solutions, including the Welch-Satterthwaite approximate degrees of freedom. Modern computing power lets us consider models with multiple “non-pooled” variances. For example, the combined analysis of an experiment repeated at multiple sites and multiple years requires at least two estimated variances: the error variance, i.e., that between replicates at the same location, and the treatment*environment variance component, i.e. the consistency of the treatment differences across environments (sites, years, or their combination). Both of these variances could be pooled or split into components. For example, a different error variance could be estimated for each site, for each combination of site and year, or for each treatment. The treatment*environment interaction could be split into separate components for treatment*year, treatment*site and treatment*site*year. In my experience, inference with pooled variances is more

sensitive to heterogeneity among treatments than heterogeneity among environments. I will share examples that illustrate this.

11:00 – 12:00 PM

“Overview of Statistics as a Scientific Discipline and Practical Implications for the Evaluation of Faculty Excellence”

William Bridges, Dept. of Mathematical Sciences, Clemson University

Walter Stroup, Dept. of Statistics, University of Nebraska - Lincoln

Nora M. Bello, Dept. of Statistics, Kansas State University

Bruce Craig, Dept. of Statistics, Purdue University

Abstract: Open discussion and solicitation of opinions from the group on the working draft of two statements targeted at outlining the role of statistical consulting and interdisciplinary collaboration in the broader scientific discipline of Statistics and practical implications for evaluation of scholarship in related faculty positions. Please refer to enclosed working documents titled “1 - Overview of the Discipline and Implications for Faculty Excellence.docx” and “2 - Position Description and Criteria for Excellence.docx”.

12:00 – 1:00 PM

Lunch

1:00 – 1:40 PM

“Multi-Trait Genomic Selection in Plant Breeding Programs”

Kelci J Miclaus, Russ D. Wolfinger, Luciano da Costa e Silva, Lauro Hose Moreira Guimaraes, Advanced Analytics R&D for the JMP Life Sciences division, SAS Institute

Abstract: Predictive models that leverage genomic variability allow crop improvement programs to make large gains in efficiency and logistical resources via virtual breeding. In nearly all breeding programs there are multiple traits of interest which can be at cross-purposes. For example, when selecting potential crosses a breeder should have the responsibility and capability to not only drive gains in yield and disease resistance, but also in vitamin content and production ease to improve global human health. In this presentation, we will focus on prediction methods such as RRBLUP (ridge regression) and GBLUP (Genomic best linear unbiased predictor) utilized in a cross-validated model comparison scheme to compare and trade-off multiple traits simultaneously. Model scoring is used to perform analytic cross-evaluation and selection based on multi-trait optimization, and then combined with progeny simulation to execute a multi-year in-silico breeding program that allows breeders to assess thousands of potential crosses that would not be feasible in the field. Corn breeding data provided by the Brazilian Agricultural Research Corporation (Embrapa) will be used as demonstration.

1:40 – 2:10 PM

“The Influence of Peak GWAS Associations on Genomic Prediction Accuracy”

Brian Rice and Alexander E. Lipka, Dept. of Crop Science, University of Illinois

Abstract: Some of the most important agronomic crop traits of interest are complex and thus governed by many genes of small effect. The statistical models typically used in a genome-wide association study (GWAS) and genomic selection (GS) quantify the contributions of genomic markers in linkage disequilibrium with these genes to trait variation. In general, the GWAS has been successful at identifying genomic regions containing markers with moderate to strong marker-trait associations. It is possible to incorporate markers tagging such GWAS signals into breeding programs through marker-assisted selection, where plants with favorable alleles at the peak GWAS signals are selected for the next cycle of breeding. In the absence of such signals, GS is typically effective at accurately predicting trait values. These two strategies have been used separately until recently, when the predictive ability of GS models that include peak associated markers from GWAS as fixed effect covariates was assessed. Theoretically, these models should be optimal for predicting traits that have several genes of large effect and many genes of smaller effect. We expand upon this work by evaluating simulated traits from diversity panels in maize using a Ridge Regression Best Linear Unbiased prediction (rrBLUP) model that included fixed effects. Upon completion of this work, we anticipate being able to rigorously quantify the ability of fixed effect covariates tagging peak GWAS signals to increase GS prediction accuracy in the rrBLUP model under a wide variety of genetic architectures and genomic backgrounds.

2:10 – 2:40 PM

“Genome Wide Association Study for Non-normally Distributed Traits: A Case Study for Stalk Lodging in Maize”

Esperanza Shenstone and Alexander E. Lipka, Dept. of Crop Science, University of Illinois

Abstract: The abundance of new genomic information available has increased the ability of statistical and computational tools to study the genetic basis of agricultural traits. As such, the genome-wide association study (GWAS), in which statistical tests of association are conducted between genome-wide marker sets and traits of interest, are one of the most predominant analyses used to dissect the genetic architecture of traits. One limitation of the statistical models typically employed in a plant GWAS is the assumption that the error terms follow a normal distribution. This project uses a novel application of statistical approaches to rigorously quantify the genomic underpinnings of a non-normally distributed trait of agronomic importance, namely stalk lodging in maize. We developed an analytical pipeline that will enable us to focus on a subset of markers in which to apply a statistically rigorous but computationally intensive model that accounts for putatively false positive signals correlated with population structure and familial relatedness. Through this methodology peak associated SNPs were identified. Future research for this project will include a simulation study to test how well this approach can identify QTL associated with binomial traits.

2:40 – 3:00 PM

Afternoon break

3:00 – 3:40 PM

"Imputation and paired samples in microRNA profiling"

John Stevens, Dept. of Mathematics and Statistics, Utah State University

Abstract: microRNAs are small non-coding RNA molecules that post-transcriptionally regulate gene expression, and can play an epigenetic role in many systems. This talk will include a summary of two statistical issues that arose in a recent microRNA profiling project where most subjects provided paired samples (from both healthy and diseased tissue). The first involves imputation of microRNA expression for subjects providing only one sample, and the second involves quantifying the benefit of various percentages of subjects with paired samples.

3:40 – 4:20 PM

“Metabolomics: Challenges and Success”

Kathleen M Yeater, USDA-ARS-NRRC-PA

Abstract: Metabolomics is defined as the comprehensive analysis of all small-molecule compounds that can be found in biological samples, such as cells, tissues or biofluids (Fiehn, 2002). Interest in this research field stems from research goals of identifying biomarkers of disease, or quantifying biochemical phenotypes. Advances in mass spectrometry (MS) and variations of such compound separation technologies also spur the interest and growth. This presentation at NCCC-170 serves as an introduction to a popular data analysis package, MetaboAnalyst (Xia & Wishart, 2011). The goal is to prepare fellow statistical consultants and collaborators with the process of this technique, and corresponding analyses options and interpretations. An experimental design and analysis will be presented. The experiment consists of a two-way balanced factorial treatment structure: (Salmonella challenge or Control) and (Prophylaxis treatment or Placebo), n=10 bovine per treatment group. The biological sample from each animal is blood serum, and metabolomics assay is gas chromatography/mass spectrometry (GC/MS). Raw spectral data processing steps is performed by a third-party contractor. MetaboAnalyst 3.0 web-based platform is utilized to perform data processing steps such as data normalization and transformation, as well as data analysis such as exploratory statistical analysis, ANOVA, and partial least squares/discriminant analysis (PLS/DA). The advantages and disadvantages of this pipeline and its results will be presented and discussed.

4:20 – 5:00 PM

Group Discussion/Adjourn.

<p style="text-align: center;">NCCC-170 / USSES Joint Annual Meeting June 22 – 23, 2017 Cargill Innovation Campus, Elk River, MN</p>

<p style="text-align: center;">Business Program</p>
--

Friday, June 23, 2017

9:00 – 10:00 AM

Business Meeting

<p style="text-align: center;">Social Program</p>
--

Wednesday, June 22, 2017

5:30 – 7:00 PM

Informal social gathering: drinks, light h'dourves and grilling.

Friday, June 23, 2017

10:00 – 11:30 AM

Walking tour of Cargill Experiment Station, led by senior scientists.

11:30 AM – 12:00 PM

Boxed lunch.