NCCC-170 Annual Meeting
June 28-29, 2018
Arkansas Agricultural Experiment Station, Fayetteville AR

Technical Program

Thursday, June 28, 2018

8:00-8:30 AM

Registration / check-in / pastries, juice, coffee

8:30-8:40 AM

Welcome and Introduction:

Nathan McKinney, Assistant Director, University of Arkansas Experiment Station

8:40-9:20 AM

“Quantification of the Genomic Contribution towards Food and Energy-related Crop Traits”

Alex Lipka, Dept. of Crop Sciences, University of Illinois

Abstract: Statistical approaches for genome-wide association studies (GWASs) and genomic selection (GS) have enabled the identification of genomic loci associated with agronomically important traits while controlling for false positives and the use of genome-wide marker data to accurately predict trait values. Using these developments as starting points, the Lipka Lab at the University of Illinois is exploring the optimization of GWAS and GS approaches and their implementation into freely available software packages. In this presentation, three different examples of research projects from the Lipka Lab are presented. Collectively, these highlight the impact of genomic properties underlying a trait and species on the performance of statistical approaches for GWAS and GS.

9:20-10:00 AM

“Searching for causal networks in experimental data: a swine production application”

Nora Bello, Dept. of Statistics, Kansas State University

Abstract: Efficient agricultural production systems require integrated management of complex physiological mechanisms. Recent developments in network methodologies can enable meaningful directional insight into the inner workings of such complex systems. Motivated by a designed experiment in swine production, we explore potential causal biological relationships between physiological outcomes in high-performing gilts and sows using structural equation models implemented in a mixed modeling framework. Data consisted of short- and long-term reproductive outcomes for 200 sows and 440 gilts arranged in a randomized block design and randomly assigned to nutritional treatments during late gestation. We implement structure-learning algorithms adapted to a hierarchical Bayesian framework to search for and quantify causal links between physiological traits separately for gilts and sows, while recognizing the multilevel architecture of the data given by the experimental design. Using a modified Jackknife resampling approach, we evaluate stability of the learned network structures and make power considerations for network inference. Results indicate distinctly heterogeneous networks for gilts and sows, consistent with differences in their physiological mechanisms. These findings have practical implications for differential management of gilts and sows to improve efficiency of swine production systems. In addition, these finding motivate further methodological extensions to structural equation models to enable specification of heterogeneous networks.

10:00-10:15 AM

Morning Break

10:15-10:55 AM

“Discrete Time Survival Analysis Applied to Experimental Data”

JungAe Lee-Bartlett, Agriculture Statistics Lab, University of Arkansas

Abstract: Time-to-event outcomes are common in agricultural sciences. For example, how long it takes until flowering is one of the critical research questions for plant scientists. Despite the popularity of survival analysis in medical studies for past decades, application in agricultural sciences has been less discussed. The main strength of this method is their ability to handle missing data over time, namely, right-censored data. Even well-designed experimental data may encounter drop-out, which can be ignored in methods such as analysis of variance (t-test) for comparing survival times for two (or more) groups. Survival models also tend to have greater statistical power to detect a significant treatment effect than methods for binary response such as logistic regression. The goal of this study is to review basic concepts of survival analysis, importantly discuss the benefit of this method when it comes to agricultural research applications. Examples vary such as time-to-damage of seed quality, time-to-clean of water, time-to-death of animals or plants. This paper demonstrates the advantage of survival analysis on experimental data through both simulation and real data studies.

10:55-11:35 AM

“Pseudo likelihood or Quadrature? What We Thought We Knew, What We Think We Know, and What We Are Still Trying to Figure Out”

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

Abstract: Two predominant computing methods for generalized linear mixed models (GLMMs) are linearization – e.g. pseudo likelihood (PL) and penalized quasi likelihood (PQL) – and integral approximation – e.g. Gauss-Hermite quadrature and Laplace. The primary GLMM package in R, LME4, uses a one-point quadrature algorithm. R also has a PQL package. In SAS®, PROC GLIMMIX was originally developed using the PL algorithm. Laplace and quadrature options were added in the 2008 release. The choice of methods presents a dilemma for GLMM users: which approximation for GLMM estimation and inference should one use, and why? Linearization methods are more versatile and able to handle both conditional and marginal GLMMs. On the other hand, GLMM software documentation and the literature on which it is based tend to focus on the limitations of linearization. Stroup (2013) reiterated this theme in his GLMM textbook, featuring examples of bias in estimates from PL. As a result, a “conventional wisdom” has arisen that integral approximation – quadrature when possible – is always best. However, despite the 2013 copyright, Stroup’s textbook was written in 2011. Meanwhile, experience with GLMMs and research about its small sample behavior has been on-going. Much “conventional wisdom” circa 2011 turns out not to be true. Above all, it is clear that there is no one-size-fits-all best method. The purpose of this presentation is to provide an updated look at what we now know about quadrature and PL, and to offer a “30,000 foot view” of some general operating principles for making an informed choice between the two. This presentation will included updates based on feedback and discussion at the 2018 Conference on Applied Statistics in Agriculture.

11:35 AM – 12:05 PM

"A Bayesian Semi-Parametric Mixed Beta Regression Model for Disease Severity in Plants"

Raul Macchiavelli, Dept. de Ciencias Agroambientales, University of Puerto Rico at Mayagüez

Abstract: Severity progress curves are used in plant disease epidemiology to describe temporal changes in the proportion of plant material compromised by the disease. For diseases with leaf symptoms, typically the damage of several leaves is assessed on each leaf on a particular scale and then averaged to a severity index (SI). The SI is often expressed in a 0-1 scale, which naturally leads to a beta distribution. In this paper we propose a Bayesian semiparametric beta regression to model the progress of disease severity. The model incorporates splines to estimate the population-average and plant-specific curves; additional terms related to the experiment design can be also included. One of the advantages of the proposed model is that it facilitates the comparison of curves between treatments across time. We applied the proposed model to Black Sigatoka disease on banana crop data from Isabela, Puerto Rico. The MCMC scheme of the proposed model was implemented in JAGS via the R2jags package. The interpretation of the analyses and the implications for the management of this disease are presented and discussed.

12:05-1:05 PM

Lunch

1:05-1:45 PM

"Using genetic relationships to improve the design and analysis of animal science studies"

Rob Tempelman, Dept. of Animal Science, Michigan State University

Abstract: It is well established that if identifiable blocking factors account for a substantial proportion of the variability for key traits of experimental interest, then randomly assigning animals to treatments within blocks should increase statistical power. In fact, block designs also generally lead to inferences that are more robust and reproducible provided that the blocks chosen for the study are widely variable and representative of the intended target population. For moderately to highly heritable traits, blocking on families should be effective and relatively straightforward to conduct for litter-bearing species such as pigs compared to, say, cattle for example. It seems then that genetic or genomic relationships between animals should be taken into consideration when blocking for treatments in dairy or beef cattle studies. We statistically assess the benefits of blocking in traditional arrangements of large half sib or full sib families as functions of heritabilities, effect sizes, and number of families. However, recognizing that population structures may be far more complex than large sib families for cattle research, we also assess the benefits of blocking based on general pedigree and/or genomic relationship matrices as well. This blocking or clustering can be based on principal component analyses, for example, which is routinely used in quantitative genomics to identify population structure. As with traditional blocking factors, genetic effects can be readily modeled as random effects within a mixed effects model. Power analyses based on mixed effects modeling is reviewed and extended to account for more general population genetic structures compared to classical block designs. We also discuss how degrees of freedom (i.e., true biological replication) for such tests might be more appropriately inferred, particularly when genetic or family effects are partially confounded with or nested within treatments. The implications for multi-pen and multi-herd studies when the experimental unit is pen or herd are discussed in the context of the degree of genetic connectedness between pens or herds. The implications of genotype by environment and/or genotype by treatment interaction on the design of animal studies are addressed as well. Properly accounting for genetic effects, particularly for moderately to highly

heritable traits, should improve research reproducibility and facilitate a better assessment of the potential for precision management of livestock based on their genotypes and/or pedigrees.

1:45-2:25 PM

“Multi-treatment (“network”) meta-analysis in agriculture”

Laurence (Larry) Madden, Dept. of Plant Pathology, The Ohio State University

Abstract: Meta-analysis, the methodology for analyzing the results from multiple studies, has grown tremendously in popularity since being first proposed by Smith and Glass in 1977. Although most meta-analyses involve a single effect size from each study (e.g., a mean difference for two treatments or a log-odds ratio), there are often multiple treatments of interest across the network of studies. Multi-treatment or network meta-analysis (NMA) can be used for simultaneously analyzing the results from all the treatments simultaneously. With this approach, correlations of treatment effects are automatically taken into account (when an appropriate model is used), and more studies may be included in the analysis because individual studies need not contain all of the treatments of interest. In fact, NMA is typically performed for sparse study-by-treatment classifications, allowing for the combination of direct and indirect evidence of treatment effects.

NMA can be based on contrasts with a baseline treatment from each study or directly on treatment arms from each study, with the estimation of contrasts performed after the model fit. The contrast-based approach is more popular, overall, especially in medical research, thanks to the statistical work and advocacy by Lu, Ades, and colleagues. Piepho, Williams and Madden showed that the results are very similar for contrast- and arm-based methods, and equivalent under some circumstances, if the appropriate mixed model is chosen. Equivalence requires, among other things, the use of a fixed main effect of study (trial) in the model. Arm-based methods are much easier to perform with standard mixed-model software, and are straight-forward to expand for incorporation of effects of moderator-variables (study-level covariates) on the response variable.

In the plant and agricultural sciences, arm-based NMA is most common. Original observations are usually not available, so the analysis is almost always based on the summary results (e.g., means) for each treatment in each study (with weights based on the within-study variances). The most extensive use of NMA probably has been in the estimation of the effects of chemical treatments (fungicides) in controlling the most economically important disease of wheat in the world, Fusarium head blight. There are now over 300 studies in the database, with over 25 different treatments, with response variables for disease severity, toxin concentration in harvested wheat grain, and yield. The mixed-model arm-based NMA is demonstrated for this dataset, and methods are proposed to determine if treatment effects are stable over the 19 years of the investigation.

2:25-2:45 PM

“Bayesian analysis of partial cladograms resulting from free-sorting tasks”

Bruce Craig, Arman Sabbaghi, and Mark Ward, Dept. of Statistics, Purdue University

Abstract: The free-sorting task is increasingly being used to compare the sensory qualities (e.g., taste, smell) of food products. In this task, a participant initially sorts the products into groups based on their perceived similarities and then successively combines the two most similar groups until only two remain. These resulting cladograms are typically converted into an overall similarity matrix and analyzed using multidimensional scaling (MDS). While the relative efficiency of this task over pairwise evaluations increases with the number of food products, there is thought to be an upper limit on the number of products one can accurately sort. Thus, studies using this task have focused on 15 or fewer products.

In this paper, we propose methods to handle studies when the number of products is above this limit. We consider a design where each participant sorts partially overlapping subsets of products and propose a Bayesian modeling method to address the inferential challenge created by these partial cladograms. Our method facilitates the combination of information across product subsets for learning the underlying latent values for all products in a comprehensive manner. These latent values are then used to construct the similarity matrix for MDS. This model incorporates variability across participants and can be extended to include covariates to help explain this variability. We demonstrate the validity of this approach via simulation studies and apply it to a study involving 21 products that are combinations of different types and concentrations of astringent and bitter.

2:45-3:00 PM

Afternoon Break

3:00-3:25 PM

“Revising an Introductory Statistics Curriculum from the Ground Up – Challenges, Solutions, and Lessons Learned”

Nicholas Keuler and Jun Zhu, Dept. of Statistics, University of Wisconsin

Abstract: For the past few years, the Department of Statistics at the University of Wisconsin-Madison has been engaged in a revision of the curriculum for introductory statistics courses targeted to non-statistics majors. We will briefly discuss some of the issues faced, solutions attempted, and lessons learned during the process, touching specifically on our chosen learning objectives, assessment, and course structures, as well as the assignment, management, and evaluation of instructors and teaching assistants. We hope that this generates a discussion among the group about how introductory statistics courses are taught at other universities, and about best teaching practices for such courses.

3:25-3:50 PM

Group discussion

(based on concerns/discussions from Matt Kramer via Walt Stroup)

Simple summary: When researchers report study results, they often want to visualize group means with ± (2) SE (or SEM) bars, and use their overlap (or not) to summarize significant differences between group means. However, in mixed or multi-factor models those error bars (and their overlap) can be easily misinterpreted. What meaningful (and easily interpretable) alternatives exist (or can be created) to visualize differences between group means? (We’re not looking at the often awkward diffogram.)

3:50-4:25 PM

Group discussion

(based on initial discussions at the Conference on Applied Statistics in Agriculture)

Simple summary: Are current journals (JABES in particular) meeting our needs and expectations, and can we as a group put together a statement for JABES to summarize our concerns and possible solutions?

4:25-4:50 PM

Business Meeting