

February 10: **Boris Shmagin** (Adjunct Professor, Ag Bio Engineering Dept.)

*Looking for STEM Education in the Big Data*

Abstract: The joint consideration of two popular topics (“buzzwords”) “STEM Education” and “Big Data” creates a synergy which helps to better understand each one. The goal and destination of completed consideration is a concept of sustainable STEM education for state of South Dakota. To achieve progress in the field of education, one needs to rely on scientific approach. The logic of the theory of communication might be a good point to start analyzing the components of education as follow: creation of content, formulation and design of the message with presence and interaction of components from different media (story, text, graphics, pictures, animations, video), communication of the message, control of the message acceptance and “digestion” (feedback communication) and the use of language and images as the main media. An example system in this case contains the professor as a transmitter (sender) and the student as a receiver.

The presentation contains these topics:

*An Introduction  
(Specialty of this particular seminar; Education and STEM; The new things and the old ones).

*Big Data – a Phenomenon  
(The “Big” – new part; The “Data” – old part; Taken together they are an unbelievable extension of the scientific approach).

*Science as an Evolutional Development for Human Species  
(The statement, hierarchy and rituals; How does it work? Scientist as a personality).

*The Language and an Image (Video) - Few Functions as a Media  
(Natural and formal languages; An image in science and mental image; Language and image; Media to describe the flooding in Aberdeen 2008; The presentation and the message).

*Education as Communication  
(The initial model of communication [C. Shannon, 1947]; Teaching [M. Montessori]; Testing).

*Scientific Education  
(Education vs. training; Education – self-motivational and hierarchical in context, procedures and achievements; Education throughout life).

*The Proposition of the Systemic Concept and Basics of STEM Education for SD  
(Individual and institutional; Local in implication - regional in development, execution and integration; Multidisciplinary, multilevel, self-organized, sustainable).

http://www.researchgate.net/publication/260107163_Looking_for_STEM_Education_in_the_Big_Data?ev=prf_pub

February 24: **Jixiang Wu** (Assistant Professor, Quantitative/Statistical Genetics)

*minque: An R Package for Linear Mixed Model Analysis*

Abstract: Generalization of linear mixed models is an important way to help data analysis with various data structures. In this presentation, first I will briefly introduce the generalization of linear mixed model system and the commonly approaches that used for analysis and statistical tests. Then, I will introduce the potential use of an R package that I recently released to the R-CRAN community. Several real data sets will be employed to demonstrate the use of this R package so the user can follow these examples to conduct their own data analyses.

March 31: **Nels Leonard** (MS student, Statistics)

*Predicting Risk of Diabetes using NHANES Complex Survey Data*

Abstract: An important part of statistics is gathering data to make inferences on a population. However, a census is often expensive if not impossible. Surveys are a popular option for obtaining a representative sample, but they introduce sampling weights. These sampling weights require a different approach to building models, and will be explored through the use of Survey Logistic Regression. Survey Design such as Stratified, Cluster,
and Multi-Stage sampling will also be discussed. Implementation of these designs will be seen from the National Health and Nutrition Examination Survey, one of the largest survey programs in the US. This data will be used to predict the risk of Diabetes as an example of Survey Logistic Regression.

April 7: Yi Xu (MS student, Plant Science)

Genetic Association Mapping: Missing Markers, Epistatic Effects, and Applications

Abstract: Association mapping has been widely used to detect desirable genetic markers associated with traits of interest for plant and animal improvement. Missing marker data are a common and challenging issue in association mapping studies, especially as the number of markers used for these studies is large. On the other hand, selection of several a set of DNA markers with potential epistasis associated with target traits will greatly help plant improvement via a marker assisted selection approach. In this study, we first proposed a linkage based imputation method for missing marker data given available linkage information and then integrated a MDR (multifactor dimensionality reduction) method with a forward variable selection approach. We applied these methods to determine SNP (single nucleotide polymorphism) markers with potential epistasis associated with agronomic traits in barley. Also, two computer program R packages were developed for these proposed methods. The detailed results will be reported.

April 14: Eric Weber (Professor of Mathematics, Iowa State University)

Fourier Uncertainty Principles

Abstract: The "Heisenberg Uncertainty Principle" is widely known among physicists as a roadblock—it limits the amount of information one can observe about the position and momentum of a subatomic particle. The Uncertainty Principle is in fact a mathematical theorem, and one which is not a hindrance but in fact beneficial. We will look at the Uncertainty Principle in several different guises, and see how we can turn it around and use it to our advantage.

April 28: Joseph Robertson (MS student, Statistics)

Forecasting Seasonal Temperature and Precipitation Outlooks Using Neural Networking and Historical Analog Strength Models in South Dakota and Neighboring Climate Divisions

Abstract: The Climate Prediction Center (CPC) issues seasonal 30-60-90 day forecasts in the United States and are used for planning purposes on a nation-wide scale. Categorical forecasting using the concept of equal chance (EC) determines whether a forecast is above normal (A) or below normal (B) for future monthly outlooks. CPC forecasting is too broad in scope for local forecasting systems. Thus, the need for more effective small scale local forecasting is the subject of this analysis. Specific planning of accurate timelines relative to weather forecasting provides the agricultural sector in South Dakota a distinct advantage should this modeling system provide better long term forecasting.

The purpose of this study is to predict and compare a neural and analog network model relative to the CPC seasonal forecasts. The tools generated from the model are referenced against historical sixty-year half standard deviation values to determine categorical forecasts. The results are paired with actual observed categorical forecasts to populate a 3x3 contingency table for each climate division in the study.

Measures of the skill of forecasting are initially measured using the Heidke skill score. Furthermore, due to the ordinal nature of the forecast variables, the marginal probability scores are weighted giving penalties to forecasts that are farther away from the actual observed values to test the validity of the prediction measures.

As of this writing, this project is in the final stages of testing the predicted values and the resulting Heidke skill scores to verify each climate division score is not simply equal or less than a random chance. The initial results look promising as there are many climate divisions outperforming the CPC.