**Statistical Design Consulting**

**SEMESTER REPORT**

**Summer 2024**

**Client:** Akhere Olenloa  **File Number:** 24-004

**Department:** Agricultural and Biological Engineering **Major Prof:** Dr. Klein Ileleji

**Consultant:** Sumeeth Guda **Initial Meeting Date:** 01/30/24

**Meeting Attendees:** Sumeeth Guda, Akhere Olenloa, Dr. Klein Ileleji, Dr. Bruce Craig

**Statement of Problem:** To investigate the challenges that grain farmers and grain elevators face regarding adopting grain monitoring technology. Additionally, to develop a predictive model for the adoption of the technology.

**Goal of This Project:** Ph.D Dissertation, Journal Article

**Background:**

The purpose of the client’s study is to understand the state of adoption of grain storage monitoring technologies. Evidence suggests that there is a very low adoption rate despite advancements in technology. The problem the client wants to address is why only 10% of on-farm storage owned by farmers use a form of monitoring technology, and 30% of off-farm storage owned by grain elevators use a form of monitoring technology.

The client created a survey to send to both farmers and grain elevators to determine what factors ultimately influenced the farmers or grain elevators to adopt or not adopt the grain monitoring technology. Within their survey they had 5 key areas:

1. The first section is collecting demographic information about the participants. Collecting the education level, gender, and region the participants are located in.
2. The second section is about the characteristics of the grain storage facilities that the farmers or grain elevators have. It asks about their storage capacity, the type of grain collected, storage period, business period, and if there are already grain monitoring technologies in place.
3. The third section is about the technological features and drivers for adoption of the grain monitoring technology, assuming the participant already utilizes grain monitoring technology. Specifically what company their technology is from, what factors are being monitored (Humidity, CO2, Spoilage, Insects, Temperature, etc.), the frequency of the monitoring, and the use of the factor within the grain management.
4. The fourth section collects data regarding the participants’ perceptions and benefits of adopting grain storage management technologies.
5. The fifth section is asking about the challenges and constraints with respect to the adoption or use of grain storage technology.

From these 5 survey areas, the client wanted to use the data collected to answer the following research questions:

1. Is there an association of the demographic information (ex. age and gender), grain storage duration, grain storage capacity and location of grain storage on the adoption of stored grain monitoring technologies among grain elevators and farmers?
2. Do grain elevators and farmers’ experiences with managing stored grain influence their adoption of stored grain monitoring technologies?
3. How do grain elevators and farmers’ perceptions of stored grain monitoring technologies influence their adoption of stored grain monitoring technologies?

The ultimate end goal of this survey was to create a regression model to predict the adoption of grain monitoring technologies and to determine what factors are significant in the adoption.

**Progress During Current Semester:**

In spring 2024, the client and consultant worked together to refine the survey questions, and develop techniques to do logistic regression analysis on the survey results. The response variable is: Did the farmers adopt the grain management technology solutions (yes, no). Starting in summer 2024, the client finalized the survey and used some of the techniques the consultant recommended on both a pilot dataset containing 50 points, and with the approximately 30 gathered survey results (as of June 2024). The client wanted help with analyzing the survey results as the models produced statistically insignificant results with high error, lack of fit, and deviance issues.

Initially the consultant and client tried to discuss the results and see if there was any way to analyze the results of the experiments without reducing the dimension of the model or introducing bias. However, after the client showed the consultant a publication from their field about a similar study. Within that article the authors introduced bias into their categorical select all that apply questions by making it into a binary variable through identifying 1 specific category and pooling the remaining into another category.

The consultant had to have a discussion with the client that given their current situation with getting invalid results from the model and the dataset. The 2 main reasons for this problem were that he had a high dimension model with some of the categorical predictors having between 1-12 levels on top of having approximately 120 predictor variables total. Alongside having a relatively low survey response size of about 30 respondents, hence leading to an undersaturated model design which lead to the lack of fit issues. The consultant had to tell the client that to address these issues, they would need to introduce bias into their model through doing a method similar to the journal authors by transforming some of the categorical predictors into binary predictors, hence reducing the model dimension. The client was precautioned that although this could remedy the lack of fit issues, ultimately given that there are 120 predictor variables and limited data points, this approach might not completely solve the problem, but it is an improvement. The key factor is that they continue to collect data from more people to increase their degrees of freedom and improve their fit. The consultant instructed the client to determine what predictors they wanted to transform into binary variables and create a document containing the descriptions of the labeling.

Although not talked about in great detail, ultimately some machine learning approaches for dimension reduction and prediction were explained although not explored deeply since the Statistical Consulting Service only does applied statistics approaches. One approach was about support vector machines (SVM) to do classification for the predictors. And the second approaches were ridge and LASSO regression. The consultant shared resources with the client in case they were interested in exploring these approaches.

**Current Status: Continuing.**