Ninth Annual
Red River Valley
Statistical Conference

North Dakota State University
Department of Statistics

Wednesday, May 1, 2019

NDSU Memorial Union

Ninth Annual Red River Valley Statistical Conference

May 1, 2019

NDSU Memorial Union

Session 1A   **Chair: Rhonda Magel**   **Location: Badlands**

9:00-9:15	Ernest Atta-Asiamah	Distributed inference for degenerate U-statistics
9:15-9:30	Mamfe Osafo	Nonparametric Test for Nondecreasing Order Alternatives in a Randomized Complete Block and Balanced Incompletely Block Mixed Design
9:30-9:45	Badr Alnssyan	Proposed Tests for the Nondecreasing Alternative in a Mixed Design Based on the Jonckheere-Terpstra
9:45-10:00	Abdelaziz Alsubie	Proposed Nonparametric Tests for The Simple Tree Alternative For Location And Scale Testing
10:00-10:15	Tiwei Zhu	Proposed Non-parametric Tests for Equality of Location and Scale against Ordered Alternatives

Session 1B   **Chair: Gang Shen**   **Location: Arikara**

9:00-9:15	Lucas Young	Type I Error Assessment and Power Comparison of ANOVA and Zero-Inflated Methods On Zero-Inflated Data
9:15-9:30	Yehong Nan	Empirical Study of Two Hypothesis Test Methods for Community Structure in Networks
9:30-9:45	Minglian Lin	Frost Depth Prediction
9:45-10:00	Ying Lin	Genomic selection in breeding using statistical models
10:00-10:15	Xiyuan Liu	Conditional Random Field and Lasso

Keynote Speech

Session 2   **Chair: Mingao Yuan**   **Location: Badlands**

10:30-11:30	Ruiqi Liu* Ben Boukai Zuofeng Shang	Statistical Inference on the Partially Linear Panel Model under Unobservable Linearity

Session 3   **Poster Session**   **Location: Arikara**

11:30-1:00	Poster Presentations	Refreshments

Session 4A   **Chair: Ron Degges**   **Location: Badlands**

1:00-1:15	Samah Althubaiti	Proposed Nonparametric Tests for the Umbrella Alternative in a Mixed Design
1:15-1:30	Qian Wen	REML Estimation of Variance Parameter in Linear Mixed Models with EM Algorithm
1:30-1:45	Hassan Rashed Alsuhabi	Proposed Modified Nonparametric Tests for the Umbrella Alternative in a Mixed Design for Both Known Peak and Unknown Peak
1:45-2:00	Xiaoxue Gu	A simulation study of skewed repeated measures data using the MIXED model
2:00-2:15	Boampong Adu Asare	Nonparametric tests for umbrella alternative in a mixed design for a known peak

Session 4B   **Chair: Bong-Jin Choi**   **Location: Arikara**

1:00-1:15	Lizzy Jepng'etich	Survival Analysis and Modelling for Cancer Victims
1:15-1:30	Madison Matthiason	Survival Analysis for Leukemia across Children, Adolescents, and Young Adults
1:30-1:45	M.G. Mostofa	Prediction of Lung Cancer Survival using Machine Learning and Statistical
1:45-2:00	Lizzy Jepng'etich Michael Heim Josh Thompson	Logistic Regression and Survival Analysis for Veteran Lung Cancer
2:00-2:15	Ibrahim Soumare	Time Series Modeling of Teens Birth Rate
2:15-2:30	Jerry Dogbey-Gakpetor Hyeongjin Jeon Juwon Choi	Text Mining and Social Network Analysis

Abstracts for Speakers

Author: Ruiqi Liu (Keynote Speaker), Ben Boukai, and Zuofeng Shang

Title: Statistical Inference on the Partially Linear Panel Model under Unobservable Linearity

A new statistical procedure, based on a modified spline basis, is proposed to identify the linear components in the panel data model with fixed effects. Under some mild assumptions, the proposed procedure is shown to consistently estimate the underlying regression function, correctly select the linear components, and effectively conduct the statistical inference. When compared to existing methods for detection of linearity in panel model, our approach is demonstrated to be theoretically justified as well as practically convenient. We provide a computational algorithm which implements the proposed procedure along with a path-based solution method for linearity detection, which avoids the burden of selecting the tuning parameter for the penalty term. Monte Carlo simulations are conducted to examine the finite sample performance of our proposed procedure with detailed findings that confirm our theoretical results in the paper. Application to Aggregate Production Data also illustrates the necessity for detecting linearity in the partially linear panel model.

Author: Badr Alnssyan

Title: Proposed Tests for the Nondecreasing Alternative in a Mixed Design Based on the Jonckheere-Terpstra

Nonparametric tests are commonly used in the field of statistics due to their robustness when the underlying assumptions are violated. In this research, eight nonparametric test statistics to test for nondecreasing means were proposed for mixed design consisting of a combination of completely randomized design, CRD, and randomized complete block design, RCBD. A Monte Carlo simulation study was conducted using SAS to estimate the powers of the proposed tests for a variety of location shifts among 3, 4, and 5 populations. The powers were estimated based on 5000 iterations using different underlying distributions. For every distribution, 16, 32 and 40 blocks were sampled from the randomized complete block design portion and equal sample sizes were used for the completely randomized design portion. The sample sizes for the CRD were selected so that the sample sizes were 1/8, 1/4 and 1/2 the number of blocks considered.

Author: Abdelaziz Alsubie

Title: Proposed Nonparametric Tests for The Simple Tree Alternative For Location And Scale Testing

Location- scale problem arise in many situations like climate dynamics, bioinformatics, medicine, and finance. In this presentation, we will propose several nonparametric tests statistics when we have a

simple tree design for detecting location and scale changes. A simulation study was conducted to compare the proposed test statistics under different conditions. The proposed test statistics are compared to Lepage's test when the number of population is equal to two.

Author: Hassan Rashed Alsuhabi

Title: Proposed Modified Nonparametric Tests for the Umbrella Alternative in a Mixed Design for Both Known Peak and Unknown Peak

This research proposes tests for testing the umbrella alternative when the data are a mixture of a randomized complete block and a completely randomized design. The need for an umbrella alternative raises in many aspects. For instance, in evaluating the reaction to drug dosage, the reaction might be increased as the level of drug dosage increases where after exceeding the optimal dosage a downturn may occur. A test statistic used for umbrella alternative was proposed by Mack and Wolfe (1981) using a completely randomized design. Moreover, an extension of the Mack-Wolfe test for the randomized complete block design was proposed by Kim and Kim (1992) when the blocking factor is introduced. This study proposes modified nonparametric tests for mixed design data with k treatments when the peak is known and unknown. The data are a mixture of RCBD and a CRD.

A Monte Carlo simulation is conducted to compare the power of the proposed test statistics to each other and with some of the existing tests in terms of the type I error and power. We consider the simulation from normal, exponential and t distributions with 3 degrees of freedom. For each distribution, equal sample sizes for the CRD portion are selected so that the sample size, n, is 6, 10, 16 and 20. The number of blocks for the RCBD considered to be half, equal and twice the sample size for each treatment. Furthermore, a variety of location parameter configurations are considered for three, four and five populations. Keywords: Mack-Wolfe test; Kim-Kim test; Modified test; Mixed design; Monte Carlo; Umbrella alternatives.

Author: Samah Althubaiti

Title: Proposed Nonparametric Tests for the Umbrella Alternative in a Mixed Design

Several of nonparametric tests are proposed for mixed design consisting of a randomized complete block design (RCBD) and a completely randomized design (CRD) under the umbrella hypothesis with known and unknown peak. A simulation study was conducted to investigate the performance of the proposed mixed design tests under many different cases such as, using different number of treatments, assuming different type of underlying distribution, changing the proportions of the numbers of blocks in the RCBD portion to the sample size in the CRD portion, and varying of configurations for the location parameter.

We consider three different proportions of the numbers of blocks in the RCBD portion to the sample size in the CRD portion. In either case, the estimated power by the first method used to proposed test statistics in this proposal is better than the second method for all different distributions. Besides that, the primary results for *four and five treatments at peak 2* indicate that:

- *For the case of distinct peak* the estimated power for the proposed test statistics in this proposal is better than the test statistics introduced by Magel et al. (2010) under the umbrella hypothesis as long as, the location parameter in left side of the umbrella hypothesis (*upside*) is **greater** than the all different location parameters in the right side of the umbrella hypothesis (*downside*) such as, (0.8 , **1.0** , 0.75 , 0.2) ; (0.75 , **0.8** , 0.6 , 0.4 , 0.2).

- *For the case of the indistinct peak with first location parameter* the estimated power for the test statistics in this proposal is still higher than the test statistics introduced by Magel et al. (2010) under the umbrella hypothesis for any configuration of the right side of the umbrella hypothesis.

There is an additional case for distinct peak for *four treatments at peak 2*; once the location parameter of the *upside* is **equal** to the first one and **greater** than the last one of *downside* like, (0.75 , **1**, 0.75 , 0.2). The estimated power of the proposed test statistics in this proposal is better than the test statistics introduced by Magel et al. (2010) under the umbrella hypothesis. However, other than the above cases the proposed test statistics by Magel et al. (2010) under the umbrella hypothesis would be preferred. Additionally, the results are similar for all distributions.

Author: Boampong Adu Asare

Title: Nonparametric tests for umbrella alternative in a mixed design for a known peak

When an assumption from a parametric test cannot be verified, the nonparametric test provides a simple way of conducting a test on populations. The motivation behind conducting a test of the hypothesis is to examine the effect of a treatment or multiple treatments against another. For example, in dose-response studies, monkeys are assigned to k groups corresponding to k doses of an experimental drug. The effect of the drug on these monkeys is likely to increase or decrease with increasing and decreasing doses. The drug effect on these monkeys may be an increasing function of dosage to a certain level and then its effect decreases with further increasing doses. An umbrella alternative, in this case, is considered the most appropriate model for these kinds of studies.

Eighteen tests statistics are being proposed for the umbrella alternative when the peak is known for a mixed design of completely randomized design (CRD), randomized complete block design (RCBD) and incomplete block design (IBD) in the first case. The second case consists of a mixed design of completely randomized design (CRD) and incomplete block design (IBD). The third case is also a mixed design of

completely randomized design (CRD) and a randomized complete block design (RCBD). We considered a combination of three tests in the first case: Mack-Wolfe (1981) for CRD, Kim, and Kim (1992) for the RCBD and Mungai (2011) for IBD. The second case is a combination of two tests: Mack-Wolfe (1981) for CRD and Mungai (2011) for IBD. The third case also is a combination of two tests: Mack-Wolfe (1981) for CRD and Kim and Kim (1992) for the RCBD.

Three situations were considered for each case: A combination without applying any modification (non-modification); a combination by applying distance modification (regular distance) and a combination by applying squared distance modification. For each combination, we considered two cases: standardized first and standardized last. The scope of this research is limited to three, four and five treatments with known peaks. In all treatments, equal sample sizes for the completely randomized design were considered and equal blocks of a randomized complete block to the incomplete blocks were considered. Studies in Monte Carlo simulation were conducted using SAS to vary the design and to estimate the test statistic powers to each other. Different proportions of incomplete blocks to complete blocks were used. Simulations from the normal, exponential and t distribution test with three degrees of freedom were also considered.

Author: Ernest Atta-Asiamah
Title: Distributed inference for degenerate U-statistics

In many hypothesis testing problems, the test statistics are degenerate U-statistics. One of the challenges in practice is the computation of U-statistics for large dataset. In this paper, we aim to reduce the computation complexity of degenerate U-statistics by the divide-and-conquer method. Specifically, we partition the full n data points into k_n even disjoint groups, compute U-statistics on each group and combine them by averaging. In this way, the running time is reduced to $O(n^m/ k_n^{m-1})$, where m is the order of the U-statistics. Besides, we study the optimal test rate of the divide-and-conquer methods. Interestingly, for degenerate U-statistics, the optimal rate is $\sqrt{n}$, faster than $\sqrt{kn}$, the optimal test rate for the non-degenerate case. The simulation and real data confirms that the proposed method has high power and faster running time.

Authors: Jerry Dogbey-Gakpetor, Hyeongjin Jeon, and Juwon Choi
Title: Text Mining and Social Network Analysis

Popularity and use of social media offer businesses another way to assess themselves as well as market their product. What customers post to a business' social media page can either be detrimental to the business or promote the business' products. Also, how well customers are connected to the social media

platform helps advertise the business and its products. This presentation **attempts** to find common words that customers of Blue Apron**, a meal kit delivery service provider,** tweet **on its social media platform (#blueapron)**, employing text mining and network analysis. **Further, we seek out how the customers are engaged and connected with Blue Apron via Twitter.**

Author: Xiaoxue Gu

Title: A simulation study of skewed repeated measures data using the MIXED model

Repeated measures data are often analyzed using a MIXED model framework which requires specification of the variance/covariance structure. The usual assumptions include multivariate normal data. We simulated data for a study consisting of 2 treatments (15 subjects per treatment group), 4 time periods (4 repeated measures per subject) with an AR(1) correlation structure. We fit 5 most common covariance structures to these data, including AR(1), Variance Components[VC], Compound Symmetry[CS], Toeplits[TOEP], and Unstructured[UN]. Their ability to control Type 1 error will be compared. The main focus of this study is to check the performance of type 1 error when simulated data under the following scenarios: 1. Non-normality of within-subject random effect; 2. Non-normality of between-subject random effect; 3. Unbalanced group size; 4. Inequality of variance of error with unbalanced group size. According to the results, we will know if the model still works well when we violate some assumptions.

Authors: Lizzy Jepng'etich, Michael Heim, and Josh Thompson

Title: Logistic Regression and Survival Analysis for Veteran Lung Cancer

United States military members have fought to protect our nation since the day it was established, but sometimes the battle does not end after returning from war. Many veterans combat war-related injuries, illnesses, or other conditions; one being lung cancer. There are many possible causes that may contribute to the apparent elevated rates of lung cancer in veterans such as the use of asbestos in supplies, weapons, or vehicles, the prevalence of smoking, or the use of chemical warfare. The Veteran's Administration conducted a lung cancer clinical trial in the 1970s to test a new treatment for lung cancer. Our project attempts to analyze this data to discover if the new treatment was more effective than the previous standard treatment at extending the life of patients, and what variables recorded in the trial were statistically significant predictors in the survival of a patient, such as the size of cancer cells, the age of the patient, or whether the patient had received prior treatment. The application of logistic regression and survival analysis methods is based on this data.

Authors: Lizzy Jepng'etich

Title: Survival Analysis and Modelling for Cancer Victims

Medical research often involves modeling for Time-to-event with censored outcome which offer more information than simply where or not an event occurred. Survival analysis should be used for the censored observation where the event was not observed during the study time. We obtained the data from the National Cancer Institute under the Surveillance, Epidemiology, and End Results (SEER) database. The SEER database includes incidence and population data associated by age, gender, race, year of diagnosis, cancer type, extend of cancer and geographic areas. The SEER program registries routinely collect data on patient demographics, primary tumor site, tumor Morphology and cancer stage at diagnosis, first course of treatment and follow-up for vital signs. In this research, we shall perform hypothesis testing to determine if there are any significant differences in survival functions for separate groups who underwent a given treatment using log-rank test. This analysis will also investigate the Cox Model and discuss more about the hazard and survival functions. **Keywords**: Regression; Survival Prediction; Kaplan Meier Estimates, Cox Model, Censoring

Author: Minglian Lin

Title: Frost Depth Prediction

Frost depth prediction is useful in many applications in several domains. In this study, data is collected from North Dakota Agricultural Weather Network Fargo station. Lasso regression is used to model the frost depth. Root mean square error is used to evaluate goodness-of-fit of the model.

Author: Ying Lin

Title: Genomic selection in breeding using statistical models

A widely used breeding method is genome selection (GS), which uses genome-wide marker coverage to predict genotypic values for quantitative traits. Genomic selection combines molecular and phenotypic data in a training population to obtain the genomic estimated breeding values of individuals in a testing population that have been genotyped but not phenotyped. The prediction of genomic breeding values is currently undertaken using multiple methods with varying degrees of complexity, computational efficiency and predictive accuracy. These popular methods comprise ridge regression (RR), ridge regression best linear unbiased prediction (RR-BLUP), the least absolute shrinkage and selection operator (lasso), elastic net. To further simplify data collection efforts and costs, we built generalized linear models without marker to predict some quality traits, and compare the accuracy of this new

approach with RR-BLUP using spring wheat breeding data. The goal of this approach is to enable the analysis of large-scale data sets to provide relatively accurate estimates of quality traits.

Author: Xiyuan Liu

Title: Conditional Random Field and Lasso

There are several big data researches applied Conditional Random Field (CRF), which is a very general and flexible to labeling observations. However, along with this flexibility, CRF has several problems. One of the most popular problems is the high-dimensional parameter problem. That is, the CRF contains too many parameters, which will cause the model to overfit the data. Besides, within these many parameters, it is possible that some of them may not have a significant influence on the model. Hence, it is very important to eliminate these parameters to avoid the risk of over fit and make the model more efficient. To solve this problem, LASSO is introduced. This presentation will introduce LASSO to improve CRF.

Author: Madison Mathiason

Title: Survival Analysis for Leukemia across Children, Adolescents, and Young Adults

The National Cancer Institute collects data on cancer patients, including diagnosis, treatments, and survival in the Surveillance, Epidemiology, and End Results Program (SEER) dataset. Using SEER's survival and sensory data I applied the Kaplan Meier Estimator to compare Male vs. Female survival rates for patients diagnosed with Leukemia in 3 age groups: children (0-15), adolescents (15-19) and young adults (19-39). I also used Cox Proportional Hazards Regression to take into account the effect of several additional variables on Leukemia survival rates including stage, tumor size, age and location. I then standardized my results in order to compare survival rates across the three age groups.

Authors: M.G. Mostofa

Title: Prediction of Lung Cancer Survival using Machine Learning and Statistical Techniques

Lung cancer is one of the leading causes of cancer death in the US. It is the second most cancer in both male and female. We will use Surveillance, Epidemiology, and End Results (SEER) large datasets for our study. Indeed, we will explore which machine learning techniques yield more predictive information about lung cancer patients. Another goal is to point out attributable variables which are more influential for lung cancer patients. For our research in this study, we would apply various supervised machine learning techniques to estimate error rate. Finally, we have the goal to enable comparison of error rate among the methods. Keywords: Decision Tree, Random Forest, Error Rate

Author: Yehong Nan

Title: Empirical Study of Two Hypothesis Test Methods for Community Structure in Networks

Many real-world data can be formulated as graphs, where a binary relation exits between nodes. These data are called network data and the analysis of these data is named the network data analysis. One of the fundamental problems in network data analysis is community detection, clustering the nodes into different groups. Many algorithms have been proposed for this purpose. However, given a network data, whether community structure exits or not is an important question. Statistically, this problem can be formulated as hypothesis testing: under the null hypothesis, there is no community structure, while under the alternative hypothesis, community structure exists. Several test procedures have been proposed in the literature. One is of the method is to use the largest eigenvalues of the scaled adjacency matrix proposed by Bickel and Sarkar(BS), which works for dense graph. Another one is the subgraph counting method proposed by Gao and Lafferty (GL), valid for sparse network. To our knowledge, for moderately sparse network, there is no test method in the literature. In this paper, firstly, we empirically study the BS or GL methods to see whether either them works for moderately sparse network; secondly, we propose a subsampling method to reduce the computation of the BS method.

Author: Mamfe Osafo

Title: Nonparametric Test for Nondecreasing Order Alternatives in a Randomized Complete Block and Balanced Incompletely Block Mixed Design

In this paper, we propose eight new nonparametric test useful for testing against nondecreasing order alternatives for a mixed design involving randomized complete block and balanced incomplete block design. Standardized first and last was calculated for Jonckheere-Terpstra, Modied Jonckheere-Tepstra, Squared Jonckheere-Terpstra and Alvo & Cabilio's test for RCBD and BIBD. Three, four and five treatments were considered with different location parameters under different scenarios. For three and four treatments, 6, 12, and 18 blocks were used for the simulation, while 10, 20, and 30 blocks were used for five treatments.

Different tests performed best under different block combinations, but overall the standardized last for Alvo outperformed the other test when the number of treatments and number of missing observations per block increases. It was also revealed that, the standardized first for MJT and Alvo had the same power throughout the simulation. The power for the standardized test for MJT and Alvo were the same for the individual designs (i.e., RCBD and BIBD). For mixed design the standardized first for MJT and Alvo were the same but different for standardized last. We can conclude that, you can apply MJT (row comparison test) and Alvo for RCBD and BIBD because they have the same power.

Authors: Ibrahim Soumare

Title: Time Series Modeling of Teens Birth Rate

Teens in the United States are far more likely to give birth than in any other industrialized country in the world. U.S. teens are two and a half times as likely to give birth as compared to teens in Canada, around four times as likely as teens in Germany or Norway, and almost 10 times as likely as teens in Switzerland. Among more developed countries, Russia has the next highest teen birth rate after the United States, but an American teenage girl is still around 25 percent more likely to give birth than her counterpart in Russia.

In this paper, we are particularly interested in understanding the trend of female teens pregnancy across each age group and race since 1960 to 2015 in the United States. Time series modeling is employed to model and forecast the teens birth rate.

Author: Qian Wen

Title: REML Estimation of Variance Parameter in Linear Mixed Models with EM Algorithm

Linear mixed model is a powerful method that can be used to analyze data that are multilevel/hierarchical, longitudinal, or correlated. For linear mixed model, there are two types of parameters needed to be estimated: fixed/random effects parameters and variance parameters. Common methods used for estimating variance parameters are maximum likelihood (ML) estimation and restricted maximum likelihood (REML) estimation. While ML estimation of variance is known as downward biased, REML corrects the bias by accounting for the loss of degrees of freedom due to estimation of fixed effects. As the REML likelihood function is a nonlinear function of variance, iterative schemes such as Newton-Raphson (NR) algorithm and Expectation-Maximization (EM) algorithm were developed to obtain the estimate of variance that maximizes the likelihood. EM algorithm is good alternative of NR algorithm as it is more robust to starting point; the variance parameters are guaranteed to remain in the parameter space and the log-likelihood does not decrease after each EM iteration. However, EM algorithm can be slow to converge. A simulation study will be used to illustrate the performance of EM algorithm.

Author: Lucas Young

Title: Type I Error Assessment and Power Comparison of ANOVA and Zero-Inflated Methods On Zero-Inflated Data

Many tests for the analysis of continuous data have the underlying assumption that the data in question follows a normal distribution (ex. ANOVA, regression, etc.). Within certain research topics, it is

common to end up with a dataset that has a disproportionately high number of zero-values but is otherwise relatively normal. These datasets are often referred to as 'zero-inflated' and their analysis can be challenging. An example of where these zero-inflated datasets arise is in plant science. We conducted a simulation study to compare the performance of zero-inflated models to a standard ANOVA model on different types of zero-inflated data. Underlying distributions, sample sizes, and percentages of zeros were variables of consideration. In this study, we conduct a Type I error assessment followed by a power comparison between the models.

Author: Tiwei Zhu

Title: Proposed Non-parametric Tests for Equality of Location and Scale against Ordered Alternatives

Ordered alternatives is an important statistical problem at life-testing experiments, drug-screening studies and so on. In our research, we will propose several non-parametric tests statistics to test equality of location and scale against ordered alternatives. A simulation study will be conducted to compare the power of the various tests under different parameter configurations and different underlying distributions.

Abstracts for Poster Presenters

Author: Emily Budke

Title: Evaluating the Impact of Temperature and Width on Rubber Band Snapping Length

Rubber bands come in all different shapes and sizes, and as such can snap at very different lengths. This experiment aims to analyze the effect of two factors on rubber bands, width and temperature. The first compares how far three different widths of rubber bands can stretch before snapping. The second will compare stretching of room temperature rubber bands against rubber bands left in the freezer overnight. This is a full factorial model with interaction to see how temperature and width affect rubber band snapping length.

Author: Casey Christianson

Title: Assessing Effects of Paper Airplane Design and Material on Distance Traveled

Have you ever taken the time to make a paper airplane, only to throw it and be disappointed? That is the reason I decided to do this experiment. Whether you are entering a paper airplane competition or just want to impress your friends, this experiment will determine how to get a paper airplane to travel farther. I will focus on two factors: design and material. I will perform 24 runs total with 4 different designs and 2 different materials of paper. Thus, each of the 8 treatments will have a sample size of 3. The response variable will be distance traveled. I want to find out whether or not both factors are significant, and I will also include the interaction effect in my model. Overall, I will implement a two-factor factorial design with two main effects and one interaction effect to assess the impact of the effects on the response variable.

Author: Jackson Denny

Title: Evaluating the Effect of Surface Type and Weight on the Bounce of a Bouncy Ball

Everyone played with bouncy balls as a kid. But did you ever wonder if the ball seemed to bounce higher on one surface compared to another? Or if the weight of the ball affected the bounce? These are the questions under consideration. We will be observing the effect on the bounce of four surface types: cement, tile, hardwood, and carpet. Furthermore, we will investigate how balls of three different masses alters the bounce of the ball. Each type of ball will be bounced on each surface with replication. This is a full factorial effect model with interaction to assess the effect of surface type on weight.

Author: Kevin Duffy

Title: Evaluating the effects of paper type on the aerial distance traveled of paper airplanes

Different types of paper may appear the same, but in reality can differ greatly in size, weight, and thickness. As such, the question can be asked, what, if any effect do these factors have upon the distance a thrown paper airplane can travel in the air. Using a two-factor factorial design, with the factors being type of paper, and the individual throwing the airplane, the distance traveled for each throw was measured and recorded. Each instance of the experiment was performed indoors to eliminate the effect of environmental factors and ultimately the findings were analyzed to determine the impact of paper type on the distance traveled.

Author: Jesus M. Elizondo

Title: The Effects of Cooking Time and Popcorn Type on Popped Corn

College students have very little free time and even less cash, as a result, it is interesting to investigate the effects of cooking time and popcorn type on unpopped corn. On the surface, popcorn is easy to make, but what is the optimal cooking time and popcorn type that will minimize the amount of unpopped corn? This experiment attempts to analyze the effects of two factors, cooking time and popcorn type, on unpopped corn. The first factor compares how 4 different cooking times affect the popcorn. The second factor compares two different popcorn types, white and yellow corn. A full factorial effect model was used with interactions to determine how different cooking times and popcorn type affect the amount of unpopped corn.

Author: Mallory Fenske

Title: Evaluating the Impact of Sales and Day of Week on Taco John's Drive Thru Speed

What makes service faster or slower when you are going through a drive thru at fast food restaurants? This experiment takes drive thru data and focuses on two factors: sales level, meaning the amount of sales for a single store monthly which will show volume, and day of the week. Additionally, the experiment uses a blocking factor of store. Data was collected from nine Taco John's, three from three different store sales levels. The response variable collected is average total drive thru time for each day of the week: Sunday through Saturday. A two-factor factorial model was used to test if there was an effect on drive thru speed depending on store level and day of week.

Author: Jamis Gruetzmacher

Title: Analysis of Poisson Surface Reconstruction with Point Cloud Data

The ability to produce three-dimensional geometry from images and video is possible through the use of surface reconstruction algorithms. However, the effectiveness of the algorithms are dependent upon input data and computational resources. In this experiment, a two-factor factorial model will be utilized to analyze the performance of the Poisson Reconstruction Algorithm. The quantity of images used as input for the reconstruction algorithm and the density of the generated point cloud from the reconstruction algorithm will be the two factors. The response variable is the reconstruction error – the distance function between the point cloud and constructed surface geometry.

Author: Turner Gunderson

Title: Analyzing the Effect of Bowling Ball Weight and Specific Lanes on Bowling Scores

Every Friday or Saturday night at NDSU, many students go to the bowling alley in the Union because it is free for weekend madness. A common debate for casual bowlers is what bowling ball to use- whether to throw a light ball very fast or a heavier ball for more force. Bowlers may have also heard of oil patterns on lanes, but may not understand what that means exactly. This completely randomized block design experiment focuses on 2 factors, the first being ball weight, and the second being the specific lane. The goal is to see whether a bowler can increase their score and impress their friends simply by picking a properly weighted ball and a favorable lane.

Author: Nils Hagen

Title: Maximizing Horizontal Distance of Nerf Blaster Projectiles through Dart Type Optimization

Over the 30-odd years of Nerf Blasters' existence, a variety of dart and gun types have been developed. It is important to have the best ammunition possible, so this experiment will test which of three styles of darts travels farthest. This experiment has a two-factor factorial design. The first factor compares how far each dart type travels. The second compares the effects of using different guns. Interaction effects will be used to see if certain dart types travel farther when fired from certain guns.

Author: Ben Hokanson

Title: Evaluating if a Golf Ball Rolls Farther than a Basketball

For those of you who have always wondered which ball rolls farther, a basketball or a golf ball, this experiment is for you. This experiment is a two-factor factorial design. The first factor is the type of ball. The next factor is which height is ideal to maximize how far the ball will roll. 10 different golf balls

will be rolled and 5 different basketballs will be rolled at each height. Then they will be compared to see which ball and which height is the best.

Author: Kristen Hurdelbrink

Title: Examination on the Significance of Milk Fat Percentage and Over-Frothing on Milk Froth Yield

Milk is known to froth up differently depending on its temperature, but what about other factors? The purpose of this experiment is to control for temperature to evaluate the possibility of any significant difference in the amount of milk froth produced across milks with different fat percentages. This experiment also considers the factor of number of times the chosen handheld manual milk frothing device is pumped. Namely, it considers the question of whether over-frothing causes a smaller milk froth yield. It is a two-factor factorial experiment to understand the full relationship between the two factors and determine what milk fat percentage and amount of pumping will produce the most milk froth.

Author: Morgan Jacoby

Title: Impact of Hand Location in Attentional Target Detection Task

With distractions all around us, focusing one's attention on something can often seem difficult. Can our own bodies be distractors? This experiment hopes to answer whether or not the location of a person's hand can affect how one performs in a common target detection task. Four factors are being investigated, hand position, response hand, target side, and validity. Hand position is defined to either be up near the computer screen or down in the subject's lap. Response hand is simply hitting the response key on a keyboard with a finger from either the right hand or the left hand. Target side indicates if the target appeared on the left side of the screen or the right. And validity looks at whether a valid cue or invalid cue as to where the target may appear has any effect on how people perform. The time it took people to respond to a new target appearing on screen was compared in this full factorial effect model to help determine what, if anything, about a person's hands will pull attention away or towards a target.

Author: Xi Ke

Title: Evaluating the decrement of used ink (mm) in writing the same article with different brands' black ballpoint pen and the type of paper

Many people use ballpoint black pen to write some calculations, reports or articles in the paper during the normal time but sometimes we feel that the black ballpoint pen run out of its ink so fast which make us need to replace it. This experiment is focusing on two factors. The first factor is the brand of the black

ballpoint pen with five levels. The second factor is the type of paper with two levels. This is a two-factor factorial design and a full factorial model was used to test the main and interaction effects to see how the brand of black ballpoint pens and type of paper affect the decrement of used ink (mm).

Author: Shubhit Kumar

Title: Evaluating the Effects of Tennis Ball and Floor Types on Energy Retained After One Bounce

Tennis balls are usually numbered 1 to 5, but is it possible that those numbers could have an impact on performance? This experiment focuses on two factors. The first factor is which ball type retains the most energy after one bounce. The second factor is which floor type allows for the highest average energy retained after one bounce. This experiment will be done using a Two-Factor Factorial Design to see if tennis ball and floor types have an effect on energy retained after one bounce.

Author: Julia McKay

Title: Testing the Impact of Charger Type on iPhone Charging Time

With phones being so heavily relied on in daily life, a sufficient battery charge is crucial. This experiment was designed to test for the most effective way to charge an iPhone. This experiment aimed to analyze the effect of two factors, charger type and outlet type. The test consisted of three types of chargers: an apple brand charger, an off-brand charger, and a wireless charger. To test a second level effect, the iPhone was charged in different outlet locations. With the phone in airplane mode, there were no interruptions in phone usage will data was collected. Under these conditions, a conclusion will be made on the most efficient way to charge an iPhone.

Author: Jared Nash

Title: Effect of political distribution and number of candidates on election results

A goal of political parties is to promote a candidate that will appeal to a majority of voters. We can evaluate a political party's success in a simulated political game. This experiment is a full factorial model that looks at two factors: a political parties' distribution relative to a left-right model and number of candidates in a primary. The observed response variable is the proportion of votes that the political party's candidate receives.

Author: Connor O'Fallon

Title: Social Context Does Not Influence Construction of Novel Faces

Social contexts can influence how observers perceive faces. For example, using an established relationship between perceived aggression and facial width-to-height ratio, Balas and Thomas (2015) found that playing a competitive game with a confederate led observers to reconstruct the confederate's face in a more aggressive configuration. In an extension of this study, we examined whether the effects of social context on face perception generalize to faces outside of the immediate social situation. 126 female undergraduates played a ball toss game competitively, cooperatively, or neutrally with another participant. A facial reconstruction task was then used to investigate face perceptions of a referee confederate and two novel faces. Overall, there was no evidence that social context influenced performance on the reconstruction task for either confederate or novel faces. These findings not only fail to support the possibility of a generalization effect, but also fail to replicate confederate effects from the previous study.

Author: Trevor Otterdahl

Title: Evaluating Strength of Glue Bond Based on Wood Type and Glue Brand

When doing any type of wood-working, the strength of wood glue is a very important factor. The integrity of a glued joint determines the fortitude of the overall structure. There are many types of wood glue on the market, but how do we know which one is strongest? This experiment is oriented on testing the strength of three of the most popular wood glue brands, Gorilla Glue, Tite Bond, and Elmer's Glue. Using a two-factor factorial experiment, the strength of the glue of the three brands was tested by gluing together 3/8-inch dowels of two different wood types: oak and poplar. The dowel segments were cut in 3-inch pieces and then randomly assigned to be glued together with another dowel of the same wood type. Hooks were then screwed into the top and bottom of the glued dowels. Then, the unit was hung by the top hook while weight was added to the bottom hook until the bond broke; the amount of weight it took to break the bond was recorded. The main, interaction, and blocking effects were then tested for their significance.

Author: Matthew Rose

Title: Evaluating the PH levels of drinking fountains on the NDSU campus

Water is an essential part of our lives, yet we don't often think about the quality of the water when we drink it. Do we consider if the water is of good quality before we lean over for a drink? This experiment focuses on two factors. The first factor is the location of the water source. The second factor is the time

of day (morning, afternoon, night) the sample is taken from. There is a blocking factor by day the observation was taken. This is a two factor factorial design with a blocking factor to determine the difference of water quality by time of day.

Author: Jared Rudolph

Title: Analyzing the Impact of Banjo String Composition on Intonation of the Instrument

Music has many different tones, textures, and styles. These attributes can be utilized using additional attachments or changing the existing parts. With stringed instruments, specifically the banjo, one common method is to change the composition of the strings used. However, with any change comes a potential to lose the fragile tuning of the instrument. This experiment aims to identify which elements of a banjo will affect the tuning of the pitches produced, using two main factors. The first factor is designated as three common alloys used to make the strings. The second factor then compares the fret positions along the neck of the banjo, with the blocking factor chosen as the four strings common to most banjos. Hence, this two-factor factorial design, with blocking, studies the impact of these factors on the tuning of the banjo, the response variable.

Author: Grant Schwieger

Title: Evaluating the Impact of Brand and Microwaving Time on the Quality of Microwave Popcorn

Popcorn is a popular snack for people of all ages, and this experiment focuses on how two factors effect the quality of microwave popcorn. In this experiment, I will be counting the number of un-popped kernels and burnt kernels after microwaving each bag, and the lower number will mean a higher quality. The two main factors I will be looking at is the effect of two different brands and 3 different lengths of time the popcorn is microwaved. This will be a two-factor factorial design with blocking, with the blocks being which boxes I got each bag of popcorn from. There will be a total sample size of 18, with 2 brands, 3 boxes of each and 3 bags in each box. There will be an interaction effect as well to see how the different microwaving times effect each brand.

Author: Reid Segal

Title: Controlling for cooking time and cooking style to ensure a soft egg yolk

Achieving the perfect egg is a difficult task- especially when you like soft yolks. The intention of this experiment was to find the best way to make an egg without hardening the yolk. The factors of interest were cooking time and style of cooking. After every trial, the yolk was cut into and its level of softness

was rated from 0% to 100%. In order to test these two variables and also measure the interaction between the two, a two factor factorial design was implemented.

Author: Caitlin Wachsmuth

Title: Evaluating the Impact of Paper Towel Brand and Water Immersion Time on Absorbency

Not all paper towels perform the same. There are brands that absorb more water and there are brands that are stronger and less likely to tear. It's important to consumers to find a product that is efficient and cost effective. This experiment will test the absorbency of paper towels using two factors, paper towel brand and time spent immersed in water. There will be three different brands and three different immersion times. The amount of water absorbed, the response variable, will be measured with each run. This is a full factorial effect model with interaction to see how water immersion affects each brand.