

North Dakota State University
Department of Statistics

First Annual
Red River Valley
Statistical Conference

Wednesday, May 4, 2011

12:00 pm

Keynote speaker: Prof. Gang Shen

A U-test to the Isotonic Change-point Problem

Location: Loftsgard 260

Refreshment Break

Location: Loftsgard 262

2:00 pm Opening remarks, Provost Craig Schnell

Location: Walster 217

Session 1: Chair: Dr. Seung Won Hyun

Location: Walster 217

2:15 pm A Nonparametric Test for the Non-decreasing Alternative in an Incomplete Block Design, Alfred Mungai Ndungu

2:30 pm A Comparison of Nonparametric Test Statistics for Nondecreasing Treatment Effects Over a Mixed Model Design, David Mathisen

2:45 pm Comparing Over-the-Counter Drug Prices and Availability Using Nonparametric Tests, Deanna Schreiber-Gregory

3:00 pm Reported energy intake by weight status, day and Estimated Energy Requirement among adults: What We Eat in America NHANES 2003-2008, Kyal Brandt*, Alicia Carriquiry, LuAnn Johnson, and Lisa Jahns

Session 2: Chair: Dr. Gang Shen

Location: Loftsgard 260

2:15 pm A Simulation Study Based on a New BIC for Detecting Change-points, Tatjana Miljkovic* and Gang Shen

2:30 pm Testing Umbrella Alternatives for Bivariate Censored Data, Ronald Carmen Degges

2:45 pm A Comparison of the Ansari-Bradley Test and the Moses Test for the Variances, Yuni Chen

3:00 pm Proposed Nonparametric Test for a Mixed Design of Independent Samples and Paired Data, Ran Fu

Refreshment Break

Location: Loftsgard 262

Session 3: Chair: Dr. Gang Shen

Location: Walster 220

3:45 pm Factors Influencing the Success of Basketball Teams in NBA League, Kishor Devkota

4:00 pm Scoring Trends: Momentum in Football, Michael Price

4:15 pm Factors Influencing C Sequestration in Northern Great Plains Grasslands, Deepti Annam*, Larry J. Cihacek, and Rhonda Magel

4:30 pm Proposed Nonparametric Test for Simple Tree Alternative in a BIBD Design, Zhuangli Wang

Session 4: Chair: Dr. Seung Won Hyun

Location: Walster 217

3:45 pm Finding the Mass Spectrum of a Sample Placed into Liquid Solution, Yunli Wang

4:00 pm Cluster Algorithm Comparison for Ellipsoidal Data, Shane Loeffler

4:15 pm MUSIC Algorithm in Signal DOA Estimation, Siqian Liu

4:30 pm Real Estate Valuation, Joshua Krueger

ABSTRACTS

Title: A U-test to the Isotonic Change-point Problem

Authors: Gang Shen

Recent years witness increasing extremities in weather patterns in various regions. A fretting question is whether they are simply natural variation or have a systematic trend. In this talk, we introduce a U-test with a flavor of the classical Mann-Whitney test, and apply it to the Argentina rainfall data and the global warming data. Our U-test statistic has an explicit asymptotic distribution for some short-range dependence data and has asymptotic power 1 even under local alternative. Compared with the existing tests, our U-test works very well for moderate size data and is more powerful in some cases of independence or short-range dependence data.

Key words: change-point; isotonic; U-statistic; mixing; stationary; nonparametric; climatic.

Title: A Nonparametric Test for the Non-decreasing Alternative in an Incomplete Block Design

Authors: Alfred Mungai Ndungu

The purpose of this paper is to present a new nonparametric test statistic for testing against ordered alternatives in a Balanced Incomplete Block Design (BIBD). This test will then be compared with the Durbin test which tests for differences between treatments in a BIBD but without regard to order. For the comparison, Monte Carlo simulations were used to generate the BIBD. Random samples were simulated from Normal Distribution, Exponential Distribution, and t distribution with 3 degrees of freedom. The number of treatments considered was three, four and five with all the possible combinations necessary for a BIBD. Small sample sizes were 20 or less and large sample sizes were 30 or more. The powers and alpha values were then estimated after 10,000 repetitions. The results of the study show that the new test proposed is more powerful than the Durbin test. Regardless of the distribution, sample size or number of treatments, the new test tended to have higher powers than the Durbin test.

Title: A Comparison of Nonparametric Test Statistics for Nondecreasing Treatment Effects Over a Mixed Model Design

Authors: David Mathisen

This study examines the testing of treatment effects over a mixed model consisting of a Randomized Complete Block Design and a Balanced Incomplete Block Design. Two test statistics are developed. Both are compared to two existing test statistics. The power of each test statistic and confidence interval for those powers is estimated through simulation studies using SAS statistical software and then compared with one another.

Title: Comparing Over-the-Counter Drug Prices and Availability Using Nonparametric Tests
Authors: Deanna Schreiber-Gregory

A study comparing over-the-counter drugs in terms of availability and prices was performed. Samples of various over-the-counter drugs were obtained from four stores using a block design. Three tests were then conducted to analyze the data obtained for this study. The first test compared differences in treatment effects using the Friedman test for a randomized complete block design. In the second test, researchers tested for the availability of drugs using the Cochran's test. A third test used a new nonparametric test for mixed design proposed by Mathisen (2011) and Ndungu (2011) and was performed to test for non-decreasing treatment effects (prices among four stores).

Title: Reported energy intake by weight status, day and Estimated Energy Requirement among adults: What We Eat in America NHANES 2003-2008
Authors: Kyal Brandt*, Alicia Carriquiry, LuAnn Johnson, and Lisa Jahns

Objective: To describe energy intake reporting by gender, weight status, and interview sequence and to compare reported intakes to the Estimated Energy Requirement at different levels of physical activity.

Methods: Energy intake was self-reported by 24-hour recall on two occasions (day 1 and day 2), approximately 10 days apart. Weight status was calculated from measured height and weight. EER was calculated for each individual based on 4 physical activity scenarios, and differences in intake – EER and % of EER were calculated using a sedentary activity level. Means and differences were derived using SAS PROC SURVEY.

Results: On average, reported day 1 energy intake was 100-300 kcal higher than day 2. Comparing intake to EER calculated using a default physical activity level of “sedentary”, we see that overweight and obese groups tend to report intakes for both day 1 and day 2 that are below the calculated EER. Healthy and underweight groups tend to report intakes that are greater than the calculated EER for both days. This could be a result of dieting in the obese and overweight groups, or an increased tendency to under report intakes as BMI increases.

Title: A Simulation Study Based on a New BIC for Detecting Change-points
Authors: Tatjana Miljkovic* and Gang Shen

A change-point is the time point when a distinct change occurs in the value of parameters of a model, which is usually unknown to us. In this project, we would like to use a new Bayesian approach, namely IBIC, recently published by Shen & Gosh (2010) for the change-point detection. The IBIC turns out to be the lower bound of the approximate marginal likelihood of a model with a change-point, while the naïve BIC which erroneously treats the discrete change-point as a smooth parameter is actually an arithmetic average of the lower and upper bound of the approximate marginal likelihood. We simulated data from Gamma and Lognormal models as well as autoregressive time series of order one. The IBIC results are compared with that by analogue naïve BIC (nBIC). Our study confirms that IBIC outperforms nBIC for the large size data with a single change-point.

1. Shen, G. & Ghosh, J. (2010), Developing a new BIC for detecting change-points, Journal of Statistical Planning and Inference, 141, 4, 1436-1447.

Title: Testing Umbrella Alternatives for Bivariate Censored Data
Authors: Ronald Carmen Degges, Ph. D.

A comparison among six weighted logrank tests for umbrella alternatives in the k sample case with bivariate censored data was conducted using Monte Carlo simulations. The six tests are a combination of ideas from Leconte, Moreau, and Lellouch (1994), Liu, Green, Wolf, and Crowley (1993) and Magel (2003). Computer simulated data sets were produced from three underlying distributions: a bivariate generalized exponential (Kundu, and Gupta (2009)); and two bivariate normal distributions with $r = 0.8$ and $r = 0.4$. The sample bivariate data taken from each of the k (3, 4, or 5) populations were first reduced to univariate data using the transformations proposed by Leconte et al. (1994). A combination of three transformations and six weighted logrank tests was considered. A simulation study was used to compare the combinations of transformation functions and weights in a variety of location shifts among the means. Estimated powers were calculated to compare the six weighted logrank tests. Various cases of bivariate means were studied to conclude which test combination would be the best test to use for any given situation.

Title: A Comparison of the Ansari-Bradley Test and the Moses Test for the Variances
Authors: Yuni Chen

This paper is aimed to compare the powers and significance levels of two well known nonparametric tests: the Ansari-Bradley test and the Moses test in both situations where the equal-median assumption is satisfied and where the equal-median assumption is violated. R-code is used to generate the random data from several distributions: the normal distribution, the exponential distribution, and the t-distribution with three degrees of freedom. The power and significance level of each test was estimated for a given situation based on 10,000 iterations. Situations with the equal samples of size 10, 20, and 30, and unequal samples of size 10 and 20, 20 and 10, and 20 and 30 were considered for a variety of different location parameter shifts. The study shows that when two location parameters are equal, generally the Ansari-Bradley test is more powerful than the Moses test regardless of the underlying distribution; when two location parameters are different, the Moses is generally preferred. The study also shows that when the underlying distribution is symmetric, the Moses test with large subset size k generally has higher power than the test with smaller k ; when the underlying distribution is not symmetric, the Moses test with larger k is more powerful for relatively small sample sizes and the Moses test with medium k has higher power for relatively large sample sizes.

Title: Proposed Nonparametric Test for a Mixed Design of Independent Samples and Paired Data

Authors: Ran Fu

This research proposes a new test for a mixed design consisting of independent samples and paired data in testing for $\mu_1 - \mu_2$, where μ is the difference between two treatments. The proposed test combines the Wilcoxon Signed-Rank test for paired data (Wilcoxon, 1945) and the Wilcoxon-Mann-Whitney test for independent samples data (Wilcoxon, 1945 and Mann-Whitney, 1947). The Dubnicka, Blair and Hettmansperger (2002) test is also a nonparametric test which can be used for the same mixed design. The proposed test is compared to the Dubnicka et al. (2002) test in terms of estimated powers. The powers are estimated based on 10,000 iterations for different combinations of the sample sizes, different distributions and various location parameter arrangements.

Title: Factors Influencing the Success of Basketball Teams in NBA League

Authors: Kishor Devkota

Teams that have highest number of wins during regular season in National basketball Association (NBA) are usually expected to reach to the finals and win the championship. However, in recent years, we have seen some team with one of the league's best winning record in regular season struggle in the playoffs. For example, Cleveland Cavaliers had the best regular season record in last two years; still they were unable to reach to the Finals in both occasions. This paper analyzes the influence of nine different performance factors in basketball in winning games during regular season and the playoffs using logistic regression techniques. Data was collected for last five NBA seasons (2005-06 to 2009-10). Field goal percentage, opponent's field goal percentage, three-point percentage, opponent's three-point percentage, free throw percentage, offensive rebound percentage, defensive rebound percentage, turnover, and offensive turnover were the factors that were examined. The results from analysis indicated that the factors that influence success of NBA teams in regular season and the playoffs may differ. Most of the factors that were significant in regular season were insignificant in predicting playoff wins. Winning percentage of a team during the regular season might not be the indicator of team's overall success in the playoffs.

Title: Scoring Trends: Momentum in Football**Authors: Michael Price**

Momentum has been proven to exist psychologically, but little evidence has been found to prove that this psychological momentum affects how humans perform. Knowing if momentum affects performance, would allow teams and athletes to play closer to their full potential, and give people a better understanding of why streaks and other rare events occur in sports. Two data sets were gathered, one that displayed the average result of all the drives in NFL football games between 1998 and 2008, and one that displayed the average result of all the drives that occurred after a turnover in NFL games during 2010. These two sets of data were compared to see if there was a significant increase in scoring after a turnover that would be a byproduct of momentum. There was a significant increase in scoring after a missed field goal and after a failed fourth down, but not after an interception and a lost fumble. A further breakdown by yard line shows that there was a significant increase in scoring after an interception and fumble lost if the drive begins on the defense's side of the field. This data suggests that momentum occurs after the defense stops the opposing team when they are about to score. Further research on this topic is needed to determine whether this trend is a sign of momentum or if the difference is due to the increased amount of scoring during recent football seasons.

Title: Factors Influencing C Sequestration in Northern Great Plains Grasslands**Authors: Deepti Annam*, Larry J. Cihacek, and Rhonda Magel**

Soil development is influenced by the five soil forming factors; parent material, climate, landscape, organisms and time. This study was designed to examine the effects of landscape and organisms (vegetation) on carbon (C) in Conservation Reserve Program (CRP), restored grasslands, and undisturbed grasslands across the northern Great Plains of the U. S. using statistical methods. The effects of vegetation, slope, and aspect on C sequestered in the surface 30 cm of the soil for 997 sites sampled across portions of Iowa, Minnesota, Montana, and North and South Dakota were evaluated. Partial F-test was used to evaluate models to determine the significance of factors and their interaction effects. For the vegetation component of these models, cool season grasses with or without legumes showed higher levels of soil organic C than warm season grasses with or without legumes or mixed cool and warm season grass regimes. When slopes were evaluated, slopes less than 3 % showed higher levels of sequestered C than slopes greater than 3 %. Southern and western aspects showed higher soil C levels than other aspects.

Title: A Proposed Nonparametric Test for Simple Tree Alternative in a BIBD Design
Authors: Zhuangli Wang

Nonparametric tests are preferred to parametric tests either when assumptions of the data distribution are violated or when information concerning the underlying population assumptions is deficient. Therefore, nonparametric tests are widely used under the weaker assumption for the underlying populations and the lower requirements for the measurement scales. A nonparametric test is presented for the simple tree alternative to test the data in a balanced incomplete block design (BIBD), where every block contains the same number of treatments, each treatment appears the same number of times, and every possible pair of treatment appears the same number of times. The details of the proposed test statistic when the null hypothesis is true are given. This includes the calculations of the mean and variance under a variety of situations. A proposed simulation study is discussed comparing the new test to the Durbin test, which is an extension from the Friedman statistic for a complete block design to a BIBD.

Title: Finding the mass spectrum of a sample placed into liquid solution
Authors: Yunli Wang

Mass spectrometry is an analytical technique commonly used for determining elemental composition in a substance sample. For this purpose, the sample is placed into some liquid solution called liquid matrix. Unfortunately, the spectrum of the sample is not observable separate from that of the solution. Thus, it is desired to distinguish the sample spectrum. The analysis is usually based on the comparison of the mixed spectrum with the one of the sole solution. Introducing the missing information about the origin of observed spectrum peaks, we obtain a classic set up for the Expectation-Maximization (EM) algorithm. We propose employing finite mixtures for modeling the spectrum of the liquid solution as well as that of the sample. A bell-shaped probability mass function obtained by discretization of the univariate Gaussian probability density function was proposed or serving as a mixture component. The E- and M- steps were derived under the proposed model. The corresponding R program is written and tested on a small but challenging simulation example. Varying the number of mixture components for the liquid matrix and sample, we find the best model according to Bayesian Information Criterion. The initialization of the EM algorithm is a difficult standalone problem that was successfully resolved for this case. We present our findings providing results from the simulation example as well as corresponding illustrations supporting our conclusions.

Title: Cluster Algorithm Comparison for Ellipsoidal Data
Authors: Shane Loeffler

This is a comparison of clustering algorithms to find the clustering algorithm that performs the best. The clustering algorithms used include Partitioning Around Medoids (PAM), K-means, Gaussian model-based clustering (Mclust), and hierarchical clustering algorithms with Ward's, single, complete, average, McQuitty's, Gower's, and centroid linkages. A mixture of Gaussian distributions was simulated with a prespecified number of dimensions (2, 5, and 10), number of clusters (4, 8, and 16), and average overlap between the clusters (0.001, 0.005, 0.01, 0.05, 0.1, and 0.2). Datasets are then simulated from the mixtures with 10, 25, and 100 points in each cluster on average. For each combination of dimensions, clusters, average overlap, and points in each cluster, the adjusted Rand index is calculated based on each clustering algorithm. This is repeated 1000 times; the clustering algorithm with the highest average adjusted Rand will be considered the best for this particular case. A t-test is used to do pairwise comparisons between the clustering algorithms to show if there is any significant difference from the best clustering algorithm for that particular case.

Title: MUSIC algorithm in signal DOA estimation
Authors: Siqian Liu

In array signal processing, a primary branch is the estimation of the direction of the arrival signal (DOA). We can always find its application in a variety of fields such as radar, sonar, geophysics, and medical imaging. By collection of the data through the radar, sensors or millimeter, the goal is to obtain the property of source though picking up the useful characteristics. Here we introduce a method called Multiple Signal Classification (MUSIC) to extract the information from the collected data.

We consider an M -element linear array of sensors and N far-field signal sources and define the $M \times 1$ vector $a(\theta_i)$ to be the complex array response for sources at direction θ_i . The outputs of the M array elements at the k th sample are arranged in an $M \times 1$ vector:

$$X(k) = AS(k) + N(k) \tag{1}$$

where $N(k)$ is noise vector, $S(k)$ is signal vector, and $A = [a_1(f) \ a_2(f) \ \dots \ a_N(f)]$, where $a_i(f) = [\exp(-j2\pi f\tau_{i1}), \exp(-j2\pi f\tau_{i2}), \dots, \exp(-j2\pi f\tau_{iM})]$ is the steering vector, $\tau_{ij} = \frac{d_{ij} \sin \theta_i}{c}$ is the phase difference between the i th and j th sensor where d_{ij} is the distance between the i th and j th sensor and c is the propagation speed of signal. We can see that the matrix A is determined by the detection frequency f and the structure of the linear array, which is predetermined during the detection. We also assume that both the signal vectors and noise vectors are stationary, zero mean random processes and they are uncorrelated with each other. Thus we can get the covariance matrix of X as

$$R = E[XX^H] = AE[SS^H]A^H + \sigma_N^2 I \triangleq AR_s A^H + \sigma_N^2 I \tag{2}$$

Do the eigen-decomposition for R and we can get the following form:

$$R = U_S \Sigma_S U_S^H + U_N \Sigma_N U_N^H \tag{3}$$

The space spanned by eigenvectors corresponding to so-called bigger eigen-values is signal space, while the one spanned by eigenvectors corresponding to smaller eigen-values is noise space.

Theoretically, the signal space and noise space are orthogonal to each other so the steer vector is also orthogonal to the noise space. Then we have:

$$\mathbf{a}^H(\theta)\mathbf{U}_N = \mathbf{0} \quad (4)$$

When the number of samples is large enough, we can get the approximation of the data covariance matrix \mathbf{R} as:

$$\mathbf{R} = \frac{1}{L} \sum_{l=1}^L \mathbf{X}\mathbf{X}^H \quad (5)$$

However, in practice, it is not easy to get the complete orthogonality between the signal space and the noise space due to the error and the interference, so we define the following spatial spectrum:

$$P = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{U}}_N\hat{\mathbf{U}}_N^H\mathbf{a}(\theta)} \quad (6)$$

Thus, we can find the estimation of θ by maximizing the P . This method is called the *Multiple Signal Classification (MUSIC)*. In general, we define a weighting vector \mathbf{W} and get the spatial spectrum as follow:

$$P_W = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{U}}_N\mathbf{W}\hat{\mathbf{U}}_N^H\mathbf{a}(\theta)} \quad (7)$$

In (7), by defining different \mathbf{W} , we can get meet different criteria. For example, if we let $\mathbf{W} = \mathbf{a}_1\mathbf{a}_1^H$, $\mathbf{a}_1 = [1, 0, \dots, 0]^T$, then we get the *minimum norm MUSIC(MNM)*; If we let the $\hat{\mathbf{U}}_N\mathbf{W}\hat{\mathbf{U}}_N^H = \mathbf{R}^{-1}$, then we get the *minimum variance MUSIC(MVM)*. As mentioned before, due to the influence of the noise and other factors, we only get an estimation of the equation (4):

$$\sigma_i = \mathbf{a}^H(\theta)\mathbf{a}_i, \quad i = N+1, N+2, \dots, M \quad (8)$$

So the log-likelihood function is:

$$-\log(f) = (M-N)\log[\mathbf{a}^H(\theta)\mathbf{T}\mathbf{a}(\theta)] + \frac{L\mathbf{a}^H(\theta)\hat{\mathbf{U}}_N\hat{\mathbf{U}}_N^H\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{T}\mathbf{a}(\theta)} \quad (10)$$

Where $\mathbf{T} = \hat{\mathbf{U}}_N\mathbf{W}_s\hat{\mathbf{U}}_N^H$ and \mathbf{W}_s is the weighted vector of the signal space. When L is large enough, the estimation error of the *MUSIC* is asymptotically Gaussian distributed with zero means. The Cramer-Rao Bound of the *MUSIC* covariance matrix is

$$\mathbf{C}_{CRB} = \frac{\sigma^2}{2L\text{Re}\{\text{HRR}^H\}}^{-1} \quad (11)$$

In Figure.1, the detection result is shown for MUSIC algorithm when there are two signals which DOAs are 80° and 90° . In Figure.2, the *MUSIC* algorithm under different criteria are compared with the CRB bound and the results showed that in the high Signal-to Noise(SNR) ratio condition, they matched quite well. However, in low SNR condition, the performance differs.

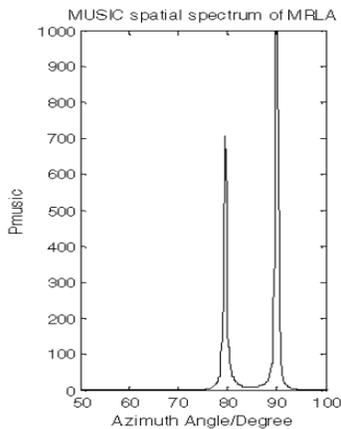


Figure 1. DOA estimation of MUSIC

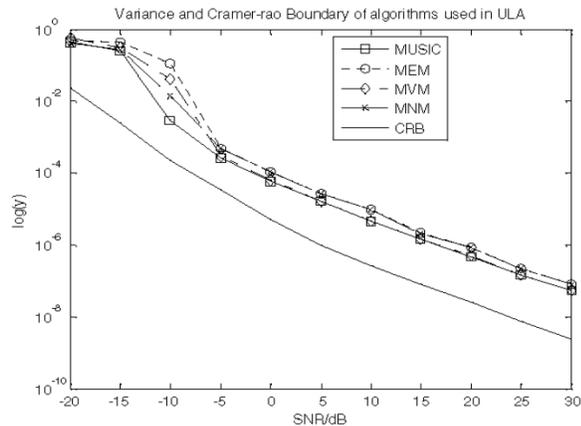


Figure 2. Statistical performance comparison

Title: Real Estate Valuation**Authors: Joshua Krueger**

A real estate appraisal is needed whenever a home is purchased or refinanced. In 2010, there were approximately 3 million home sales according to the National Association of Realtors. Fannie Mae was involved in 2.1 million refinances during the same year. Using the median sales price in 2010 (\$173,100), the total amount of loans could have exceeded 700 billion dollars. The basic valuation technique used to evaluate residential real estate is paired sales analysis. Paired sales analysis is defined as an approach to estimate values of specific property characteristics to find a value of the subject property; property sales are paired by similar property characteristics. Ideally the properties are exactly the same except for one characteristic; the difference in the sales price can be attributed to the difference in this characteristic. However, several adjustments usually must be made to paired sales analysis to isolate the effect of one characteristic. One can immediately see the problem with paired sales analysis. 700 billion dollars in loans are based on this technique. To address the problem of valuating real estate, I used regression analysis on a section of West Fargo called Eagle Run. I took 113 single family sales from the past year and created a model. The model was used to predict sales prices in that neighborhood. After creating the model, I compared sales that were not a part of the data set. The compared sales happened within a month of the most recent sale in the data set. Due to a limited time frame after the last sale date, there were only a small number of sales that could be applied to the model. The results show that regression analysis can be an additional tool to assist in the valuation process.

POSTER ABSTRACTS

Title: The Effects of Previous Trial Validity on the Gaze Cuing Effect: A Meta-Analysis

Authors: Deanna Schreiber-Gregory

A meta-analysis was performed to explore whether the validity of the previous trial affects the magnitude of the cuing effect on the current trial in gaze cuing studies. The eight selected studies for the analysis were all performed by one researcher. The hypothesis was that the facilitation for a validly-cued target (compared with an invalidly-cued target) would be greater when the target on the previous trial was validly-cued than when the target on the previous trial was invalidly-cued. The results indicated that previous trial validity does have an effect on the current trial validity and evidence of a cuing effect was well established.

Title: Classification of Time Series Analyzed Microarray Data Using Decision Tree Algorithm

Authors: Anuradha Vegi, Ushashi Chakraborty, and Anne Denton

Data of time series experiments involving microarray analysis gives details of various genes and their expression profiles of an organism at different time points of the life cycle. However, understanding microarray gene expression profiles is difficult as the data is complex. Classification can help predict the behavior of unknown genes by using already available information. In the present project a decision tree algorithm was used to classify various genes (n=6180) involved in cell cycle regulation of *Saccharomyces cerevisiae* (yeast) synchronized by alpha arrest, elutriation and arrest by cdc15 and cdc28 temperature sensitive mutant methods (alpha, cdc15, cdc28 and elu). Relatively high specificities were obtained for predicting cell cycle regulation genes of all the four gene datasets including 0.83, 0.71, 0.73 and 0.69 respectively for alpha, cdc15, cdc28 and elu methods in yeast. Overall, the algorithm has proven to have relatively high specificities, thus making it an important tool to predict cell cycle regulation genes in yeast.

Title: Diet Effect in Pigs' Weight

Authors: Giovana M. Anderson

This experiment's objective was to analyze if treatment affect the gain weight of pigs using 4 different treatments. Twenty-four gilts (n=24) were selected from the North Dakota State University research herd for inclusion in the experiment. The data was analyzed using PROC ANOVA on SAS/Lab Software. From the results we can see that the p-value (p)=0.98 is not the significant. Also, the Tukey-Kramer test was used to test the significance of each group to one another and to correct type I error. Thus, from the sample data, there is not sufficient evidence to conclude that the 4 treatment have an effect in the weight of the pigs. The pigs did not gain weight based on the treatments. All gilts were selected from a common sire and be individually penned at 8-weeks of age (± 5 days) in groups of eight per treatment.

Title: Men vs. Women Car Model Preference

Authors: Josiah Olson

A car company wants to see if there is evidence that the proportion of women intending to buy model A differed from the proportion of men intending to buy model A. In a random sample of 200 women, it was found that 62 were intending to buy model A. In a random sample of 200 men it was found that 74 were intending to buy model A. At alpha equal to 0.05, is there a significant difference in the proportions of men and women intending to buy model A?

Title: Comparison of Pace Times for Male and Female Runners

Authors: Jeney Anderson

I did a comparison of pace times for male and female runners of the Fargo Marathon. While it is an accepted fact that male runners are faster than women on the whole, I was curious as to how the averages compared. More specifically, I wanted to see if women's average pace times were slower than men's.

Data for this project was found at onlineracesresults.com via the Fargo Marathon main webpage. Once the race results were organized by gender, I then took a random sample of average pace times from both the men and women at random. Once compiled in a spreadsheet, I transferred the data to Minitab to complete my calculations.

The calculations for my data can be reviewed below.

Two- Sample T-Test and CI: Mens Pace, Womens Pace

Two-sample T for Mens Pace vs Womens Pace

	N	Mean	StDev	SE Mean
Mens Pace	43	9.25	2.04	0.31
Womens Pace	43	9.64	2.12	0.32

Difference = mu (Mens Pace) - mu (Womens Pace)

Estimate for difference: -0.397

95% lower bound for difference: -1.144

T-Test of difference = 0 (vs >): T-Value = -0.88 P-Value = 0.810 DF = 83

I came to the conclusion that there was no enough evidence to support my hypothesis that women's average pace times were slower than the men's at $\alpha = .05$.

Title: Top 20 Active NBA Free-throw Shooters
Authors: John Richardson

I gathered data on the Boston Celtics and Miami Heat basketball teams. In my data I calculated the amount of free throws made and attempted for each team. The data concluded the Boston Celtics shot a slightly better percentage from the free throw line. The Boston Celtics attempted 1,164 free throws and made 1,323 of them, for a free throw percentage of 78%. The Miami Heat attempted 2,182 free throws and made 1,687 for a free throw percentage of 77%. Although the Miami Heat attempted over a 1,000 more free throws the Celtics were able to shoot a better percentage.

Title: Military Deaths
Authors: Melissa Gillett

Over the past of couple decades, blame has been placed on terrorist groups for the high number of military deaths. A two-way ANOVA was used to examine military cause of death data from 1980-2008. This data was blocked by year and divided into categories: Accident, Hostile Action, Homicide, Illness, Self Inflicted, Terrorist, Attack, and Undetermined. This test will provide comparisons for the percentages of different causes of death found by dividing by the total number of deaths.

Title: A Comparison of Temperature Between Bismarck and Fargo
Authors: Niteen Jadhav

Paired t-statistical test was performed by using Minitab computer package. To perform this test the data for highest and lowest temperature in cities of Fargo and Bismarck was collected for the month of March 2009. The populations were assumed to have a normal distribution. The data was collected for 21 days. The hypothesis was that the mean difference in highest temperature of Fargo and Bismarck is not zero. After performing paired t-statistical test it was found that, the p-value is greater than the alpha value (0.05); the null hypothesis cannot be rejected. Hence it can be assumed that the mean difference in highest temperature of Fargo and Bismarck is zero.

Title: Comparing Percentages of Field Goals Made at Various Yard Lines
Authors: Brad Ruhanen

This one-way analysis of variance test was performed to test the claim that three or more population means (yardage) are equal. The data used for this set represents the percentage of all field goals (FG) made by thirty NFL kickers at ranges of 20-29 yards, 30-39 yards, and 40-49 yards. I tested at the 0.05 level of significance whether the differences among the three sample means could be attributed to chance. The data supports that the 20-29 yard range was the most effective with accuracy (96%), followed by the 30-39 yard range with 87%, and last was the 40-49 yard range with 72% accuracy. These results show a rejection of the null hypothesis.

Title: Testing the Top 20 Homerun Hitters of All-Time compared to Number of At Bats
Authors: Tyler Roehl

A regression analysis was conducted to determine the relationship between the number of home runs and the number of at bats for top home run hitters in major league baseball. The results of the study will be discussed.

Title: Evidence to Support 62% Sideout Percentage for NDSU Women's Volleyball Team
Authors: Kelly Lopez

I decided to conduct a hypothesis test to determine if there is evidence to support that the population mean ratio, or the Sideout Percentage of (.6192) or 61.92% results in the NDSU Women's Volleyball team winning that set of the match. When I was a collegiate athlete here playing volleyball for the Bison, I was an integral part of the team, playing the entire match as the Libero and was given the privilege and opportunity to control and dictate the backcourt. What this means in general terms, is I was the "quarterback" of the backrow, and dictated what needed to happen in Serve Receive (when other team would serve to us), as well as on Defense (when other team would attack against us). Throughout sophomore year, my coaches would emphasize that we needed to Sideout above 60% to have a good chance at winning the set. In collegiate volleyball, matches are played best 3 out of 5 sets. The first four sets, if taken that far, are to 25 points. The last set is to 15 points. During my junior and senior seasons, we were finding that if we would sideout above 62% that our chances for winning that set of the match, were extremely good. Our coaches spent a lot of time looking back at previous stats of matches, but there was never a hypothesis test done to determine if all of this technically served true. I have decided to be that person to test and possibly prove that when our NDSU Women's Volleyball Team serve receives (sideout) above 62%, that they are more likely to win that set, verses siding out at a percentage under 62.

I used a 1-Sample Z test. The data used for this set were taken from the NDSU Women's Volleyball 2010 fall statistics. I randomly chose 20 matches the women played, and of those 20 samples, 50 samples of sideout percentages were randomly obtained.

I found using a 99% confidence interval with my region of (.5720, .6664) that the .62 that I was testing, falls within the interval. Thus, I do not reject. My coaches were pretty right in their findings that siding out at or above 62% would more often than not, result in a win.

Title: The calculation compares the sodium content of two of the most common food items bread and breakfast cereal

Authors: Hiral Bhatt

The data was collected <http://www.alsosalt.com/sodiumcontent.html>

34 types of bread and 40 types of cereal were selected independently and the two population samples were compared.

Hypothesis for the test was

Ho $p_1=p_2$

Ha $p_1>p_2$

Here p_1 is the proportion of breads with sodium content greater than or equal to 170mg

And p_2 is the proportion of cereals with sodium content greater than or equal to 170mg.

This was compared at $\alpha=0.05$. This is to compare if sodium content is usually more in breads than cereal.

The test was run on using Minitab software.

After analyzing the results it was found that we cannot reject the null hypothesis as P-value was significantly high at approx 0.9.

Also the test statistic value was not greater than Z $\alpha.05$

Test statistic value -1.53 \neq .361365

Do not reject null hypothesis. This implies there is not enough evidence to state that the proportion of bread with sodium content greater than or equal to 170 mg exceeds the proportion of cereals.

Title: Grain yield of *Camelina sativa* L. at different nitrogen fertilizer rates

Authors: Itai Mutukwa

Nitrogen (N) fertilizer is important for crop growth and yield. Knowledge of yield response to fertilizer application rate is important in determining optimum application rates. An experiment was conducted to determine the effect of N fertilizer rates on *Camelina sativa* L. grain yield. Four fertilizer rates (0, 75, 150 and 300 kg N ha⁻¹) were adopted. The experiment was laid out in a randomized complete block design with four treatments, three blocks and six replicates. Grain yield was measured in kg ha⁻¹. Analysis of variance was done using Minitab. There was no significance difference in yield amongst the four nitrogen rates. However, blocking was found to be significantly effective in reducing variation.