6-2012

Regression-Based Prioritization and Data Modeling for Customized Civil Engineering Data Collection

Omar Kanaan
Western Michigan University

REGRESSION-BASED PRIORITIZATION AND DATA MODELING FOR CUSTOMIZED CIVIL ENGINEERING DATA COLLECTION

by

Omar Kanaan

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Science in Engineering (Civil)
Department of Civil and Construction Engineering
Advisor: Pingbo Tang, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2012

THE GRADUATE COLLEGE
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN

Date 5/23/2012

WE HEREBY APPROVE THE THESIS SUBMITTED BY

Omar Kanaan

ENTITLED Regression-Based Prioritization and Data Modeling for Customized Civil Engineering Data Collection

AS PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF Master of Science in Engineering (Civil)

Civil and Construction Engineering
(Department)

Pingbo Tang
Pingbo Tang, Ph.D
Thesis Committee Chair

Civil Engineering
(Program)

Jun-Seok Oh
Jun-Seok Oh, Ph.D., P.E, PTOE
Thesis Committee Member

Valerian Kwigizile, Ph.D., P.E
Thesis Committee Member

APPROVED

Dean of The Graduate College

Date June 2012

REGRESSION-BASED DATA MODELING FOR CUSTOMIZED CIVIL ENGINEERING DATA COLLECTION

Omar Kanaan, M.S.E

Western Michigan University, 2012

Civil engineers frequently face the challenges of acquiring sufficient data to satisfy the informational needs of various decision making scenarios having time, budget, and resource constraints. This thesis focuses on exploring how statistical analysis of historical data sets can be used to improve the efficiency and effectiveness of future data collection activities in transportation engineering and construction project control. Pearson's correlation, multiple regression, Akaike information criterion, and Bayesian information criterion are applied to two cases of data collection and analysis to understand the relative importance of various parameters within each case. Case I analyzes National Bridge Inventory (NBI) data to identify the cross-regional variations and the relative importance of various bridge parameters as they relate to bridge deterioration across regions in the U.S.. Substantial variations in bridge deterioration influential items are observed across different regions and rating items while some items are consistently identified as important. Case II analyzes and ranks several scanner-based, environmental, and object-based factors influencing point cloud data quality and creates a scan-error prediction model for more effective scan planning using 3D laser scanners in construction project control.

ACKNOWLEDGMENTS

I wish to express my sincere thanks and appreciation to my thesis advisor Dr. Pingbo Tang for his encouragement, guidance, insight, and support during this research and the preparation of this thesis. In addition, I would like to thank my committee members Dr. Jun-Seok Oh and Dr. Valerian Kwigizile for sharing their expertise and guiding me since the early stages of this research.

Omar Kanaan

TABLE OF CONTENTS

Table of Contents - Continued

Table of Contents - Continued

APPENDICES

~~CHAPTER~~

LIST OF TABLES

LIST OF FIGURES

List of Figures - Continued

CHAPTER I

INTRODUCTION

Background

This thesis focuses on exploring how statistical analysis of historical data sets can be used to improve the efficiency and effectiveness of future data collection activities in various civil engineering applications. Civil engineers frequently face the challenges of acquiring sufficient data to satisfy the informational needs of various decision making scenarios having time, budget, and resource constraints. Analyzing data collected in the past may reveal correlations among domain requirements, data qualities and various factors influencing the data qualities. As detailed in the following paragraphs, these correlations can assist engineers in prioritizing data collection options to minimize resource consumptions of data collection activities while ensuring the satisfaction of the domain requirements.

In generic civil engineering, the process of identifying problematic components in sophisticated systems is difficult and even impossible to achieve by simple visual observation. This difficulty has pushed civil engineers, specifically construction and transportation engineers, to analyze both historic data and data collected using state of the art technologies to better manage their projects. Recent advances in both data collection and data analysis technologies allow engineers to better identify influencing factors both positive and negative.

In the domain of infrastructure management, effective data analysis is critical for managing more than 600,000 national bridges across the United States (U.S.). Transportation engineers at both the state and federal levels have access to data collected under the National Bridge Inventory program which contains 116 data items (complete list of items found in Appendix A) relevant to bridge condition evaluations for national bridges in the U.S. (Weseman, 1995; Tang and Akinci, 2009). Systematic data analysis will allow engineers to better understand how the relative importance of different bridge components and attributes vary across regions in the U.S.. That will further allow the United States Department of Transportation (USDOT) and the Federal Highway Administration (FHWA) to better allocate funds and/or manpower to specific aspects of a project/process that justly need it.

In construction engineering, point clouds collected using 3-D laser scanners are used in a variety of applications such as indoor mapping (Tommaso et al., 2006), project control (Akinci et al., 2002), construction metrology (Cheok et al., 2001), development of as-built models for existing and new facilities (Arayici, 2007) excavations (Su et al, 2006) and resource management (Gong and Cladas, 2007). While laser scanners are capable of collecting a large number of data points with incredible accuracy under lab conditions, construction sites contain various impediments that may hinder the quality of collected data. A better understanding of the variables that affect the quality of the data collected will ensure a more cost effective data collection process in the field that will capture all needed data while reducing the interferences of the data collection activities with the construction activities.

This thesis utilized four statistical methods to examine data collection and analysis cases as they relate to construction and transportation engineers. Case I analyzed NBI data collected in 2010 and identified the cross-regional variations and the relative importance of various characteristics as they relate to bridge deterioration. Case II analyzed and ranked several factors influencing point cloud data quality in an effort to create a scan-error prediction model for guiding more effective data collection using 3D laser scanners for construction project control.

CHAPTER II

PROBLEM STATEMENT

As manually prioritizing data collection options is subjective and time consuming, it is necessary to explore automated methods for quantifying the importance of various data collection options. Generally, engineers need to conduct two types of prioritizations: 1) prioritize various data sets based on their correlations with the information of interests (e.g., bridge condition ratings) to decide which types of data sets deserve more resources and time; 2) given a data collection activity (e.g., using a laser scanner to collect 3D data), prioritize data collection parameters based on their impacts on the data qualities. The former is the first level in determining what types of data to collect, while the latter is the second level in defining specific data collection parameters to optimize the data qualities and resource consumptions. In this research, Case I was used to explore the first type of prioritization, while Case II was used to explore the second type of prioritization, as detailed in the two subsections below.

Case I – Importance Ranking of Bridge Condition Explanatory Data Items for Customized Data Collection and Bridge Management

As it is critical to rank the relative importance of various items potentially useful for bridge condition evaluation and prediction, systematic explorations in this area are still limited. Existing multi-region explorations mostly focus on comparative analyses of several factors (<10) manually identified by domain experts through interactively visualizing the data and conducting spatial statistical analyses (Chase et al. 1999; Karlaftis 2005; and Wang et al. 2010). Relying on expert knowledge and experiences may provide reliable and intuitive results, but such analyses tend to be time-consuming, ad-hoc, and subjective, making it impractical for ranking larger number of candidate explanatory data items (Glaser and Tolman 2008; Morcous 2002; and Tokdemir et al. 2000). In practice, as the number of candidate input variables could be large (>100), comprehensively identifying all critical variables for a large number of regions (e.g., 50 states in U.S.) having possibly different importance rankings of these data items would be challenging or even impractical. As a result, even though previous studies explored the technical feasibilities of using various statistical methods for data from a few regions (Bolukbasi et al. 2004; Mclure and Daniell 2002; Chase et al. 1999; Kim and Yoon 2010; Ahlborn et al. 2010; Burgueno and Li 2008; Chase et al. 1998; and Dunker and Rabbat 1995), a systematical comparative evaluation of these statistical methods on data sets from multiple geographic regions is missing.

To address the knowledge gaps pointed out above, Case I has two objectives: 1) establishing a framework using multiple statistical methods to identify significant explanatory items in large numbers of candidate explanatory data items related to bridge condition ratings; 2) using this framework to conduct comparative analyses across geographical regions and rating items across the U.S. to narrow the number of items of interest for each of the studied regions. Achieving these objectives will allow engineers to create region-specific inspection and maintenance procedures that focus on the identified items for each region.

As NBI has been a representative bridge management database used by FHWA and state DOTs for decades, this research uses NBI data collected in 2010 to demonstrate the effectiveness of this framework. Further studies can be built upon this research to guide more effective bridge data collection and bridge management in U.S.

Case II - Prioritizing Influential Factors on Laser Scanned Data Quality

Recent technological advances give engineers the ability to collect a large amount of information/data about physical objects in the field, one such technology being the laser scanner. Laser scanners create a digital three dimensional replica of the geometry scanned by collecting millions of individual points with 3-D coordinates; known as a point cloud (Yong K. Cho et al 2011).

In civil engineering, point clouds are used in a variety of applications. This range of possibilities creates an infinite amount of potential scans with variable goals and constraints for each application scenario. However, construction activities often influence the time allocated to scanning thus adding a non-physical constraint that needs to be accounted for during scan planning (time). There have been several studies addressing the need to reduce scan times without negatively influencing the quality of the collected data (B. Akinci et al. 2006; E. Latimer et al. 2004), however, these studies either focused on visibility based scan planning without in-depth considerations of the impacts of various environmental factors on the data quality, or accuracy analysis of 3D point clouds without quantifying the impacts of various factors. As a result, engineers lack quantitative awareness of the implications of their decisions on data collection parameters in the field (e.g., trade-offs between scanning distance and data densities). Multiple trade-offs among various data collection parameters make the analysis of various data collection options in the field extremely complex. To ensure that a given scan generates useful results there must be a framework that, at its core, not only defines the minimum data collection requirements i.e. domain requirements related to particular engineering applications, acceptable error, but also model the errors of point clouds to capture complex relationships between the data quality and various data collection parameters. That is, engineers should be able to tailor the data collection process to ensure that the collected data meets a set of criteria defined by the accuracy of the results. To achieve that, engineers need a methodology for estimating qualities of data given the values of data collection parameters (resolution, scanning distance).

CHAPTER III

METHODOLOGY

This thesis investigates and compares four statistical method results to identify influential variables for both cases I and II: 1) Pearson's Correlation; 2) Multiple Regression; 3) AIC based statistical model selection; and 4) BIC based statistical model selection. While these four statistical methods are at the core of the analyses, each case utilizes some other methods discussed in their respective chapters. The analyses strive to identify the strength of relationships between input and output variables; where for Case I, input variables are the bridge attributes collected from the NBI database, and the output variables are the rating items also found in the NBI database. While for Case II, input variables are the nine scan condition variables; environmental and physical conditions, and the output variables are the scan error discussed in Chapter 3.

Pearson's Correlation is a method for computing the strength of linear correlation between two variables (input, output) based on multiple pairs of values of these two variables. For each pair, Pearson's Correlation computation generates a correlation coefficient in the range of -1 to 1: 0 indicates no linear correlation between these two data items, and ±1 indicates positive/negative linear correlation between them. Larger absolute values of correlation coefficients indicate stronger linear correlations (Dalgaard 2008).

Multiple Regression assumes that the values of an output variable, Y, can be estimated as a linear function of the values of multiple input variables, x_n, with β_n as the regression coefficient of x_n, as shown in equation (1):

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon \qquad (1)$$

Y = A rating item; β_n = Regression coefficients; x_n = Explanatory items; ε = Unexplainable random noise

For each observation of Y and x_n, data from multiple observations can setup a group of linear functions in the form of (1). As the numbers of observations is much larger than the number of output variables, it is possible to estimate β_1, β_2... β_n by solving (1) through a least-square based approach. After obtaining β_n, standard hypothesis-tests developed in the domain of Statistics can be applied to the values of β_n and the estimated standard errors of β_n for establishing an "Insignificance" value (also known as p-value (Efroymson 1960) for each data item. Generally, the insignificance for a given output variable while studying a given input variable is the probability that the estimated β_n of this input variable is small enough for not rejecting the hypothesis "this input variable is linearly uncorrelated with the output variable" (Null Hypothesis in Statistical hypothesis testing). Such insignificance values, therefore, can serve as indicators of the relative importance of input variables for a given output variable: the smaller this insignificance, the less likely that the variables are linearly uncorrelated; hence the input variable is more likely to be important. In this thesis, for each of the cases, a regression analysis was conducted, and sorted input variables in ascending order of their values of insignificance. More details are in any standard Statistics textbooks (Dalgaard 2008).

AIC and BIC are two widely-adopted information criteria measuring the relative goodness of fit of statistical models (Kuha 2004).

$$\mathrm{AIC} = 2k + n \log\left(\frac{RSS}{n}\right) \qquad (2)$$

$$\mathrm{BIC} = n \log\left(\frac{RSS}{n}\right) + k\ln(n) \qquad (3)$$

k = number of parameters; n = number of observations;

RSS = residual sum of squares

$$\mathrm{RSS} = \sum_{i=1}^{n}\left(y_i - f(\hat{y}_i)\right)^2 \qquad (4)$$

y_i = i^{th} value of the variable to be predicted,

$f(\hat{y}_i)$ = predicted value of y_i

In this thesis, for a given output variable and a number of candidate input variables, several input variables (k selected variables) can be selected, and fitted with a multiple linear regression model on N observations of these selected variables (N bridges/scans) against the corresponding observations of the output variables. One such fit can generate a statistical model depicting the relationship between the K selected variables and the output variable. As K items ($0 < k < N$) can be chosen from the total number of input variables N, it is possible to generate multiple statistical models, each with a unique RSS, for multiple possible combinations of input variables. Some of these statistical models are relatively better than others in terms of reliably capturing the statistical correlations between variables (goodness of fit) while avoiding over-fitting the data (unnecessary model complexity compromising the prediction capabilities of models). AIC and BIC are both measures for statistical models considering both the goodness of fit and appropriate level of model complexity (Kuha 2004).

Therefore, to find the best model using AIC and BIC methods, each of the N explanatory variables is used as a model separately and their corresponding AIC/BIC value is calculated using equations (2) and (3). The explanatory variable with the lowest resultant AIC/BIC value then remains in the model while the remaining explanatory items (N-1) are individually input into the next set of models; forward selection. The two-variable model with the lowest AIC/BIC value then becomes the basis to create a three-variable model and so on, until the best model is found. AIC and BIC both stop adding explanatory variables at the step where adding a variable increases the AIC and BIC values. Both methods include a "penalty term" that penalizes models with more parameters k (adding 2k for AIC and k*ln(n) for BIC).

Theoretically, Pearson's Correlation studying individual variable pairs ignores the effect of other input variables when measuring the marginal effect of one particular input on the output variable. On the other hand, the other three methods take into account the effects of other input variables and thus produce a more appropriate measure of the importance of a variable. In this thesis, the influence of the combined effects and the marginal correlations of multiple input variables on the generated results are examined, aiming at obtaining theoretical insights of these statistical methods.

CHAPTER IV

CASE I: IMPORTANCE RANKING OF BRIDGE CONDITION EXPLANATORY DATA ITEMS FOR CUSTOMIZED DATA COLLECTION AND BRIDGE MANAGEMENT

Literature Review

Previous studies in the domain of bridge management show the necessity of understanding the relative importance of various data items for bridge condition rating, while such relative importance varies substantially across geographical regions. Multiple studies strived to identify critical data items having significant and reliable correlations with the current and predicted conditions of bridges (Valenzuela et al 2010; Gokey et al 2009); such data items include items describing the bridge attributes, such as bridge material and structural type (Kim and Yoon 2010), and items about the environmental conditions, such as weather and hydrologic conditions (Chase et al 1999; Karlaftis 2005). However, most previous studies focus on analyzing a single region using one or two statistical methods, while the importance of critical data items may vary substantially across regions. For instance, two studies in Virginia (Gokey et al 2009) and Illinois (Bolukbasi et al 2004) agree that the age and the Average Daily Traffic (ADT) of a bridge are important indicators of bridge conditions, while a study in North Dakota highlighted the importance of structural characteristics (materials, etc.) and the presence of water (Kim and Yoon 2010). Such cross-region variations indicate the necessity of a systematical multi-regional study (Dunker and Rabbat 1995).

Moreover, with the awareness of the relative importance of various explanatory data items for bridge conditions, bridge management agencies will be able to develop customized data collection and analysis plans to better allocate the limited resources and acquire the most useful information for given geographical regions.

Methodology

Figure 1 shows an overview of the research methodology. This exploration has three main stages: 1) NBI data preparation; 2) Statistical analysis; and 3) Comparative analyses. Stage 1 focuses on identifying and preparing data sets that will be analyzed by the statistical methods used in stage 2. In this stage, rating items identified as NBI data items that are rating the bridge conditions, similarly explanatory items are defined as NBI data items that can serve as candidate data items for explaining such condition ratings. After identifying these two categories of NBI data items, five regions across the U.S. are selected to cover a variety of geographic environments. As some data entries in NBI have missing or non-valid attribute values (e.g., 'NA' meaning "Not Applicable" for some attributes), which might mislead the statistical analysis, the last step in stage 1 is data cleaning on the selected NBI data sets. Subsections "NBI Data Used in this Research" and "Data Cleaning" below will present details of stage 1.

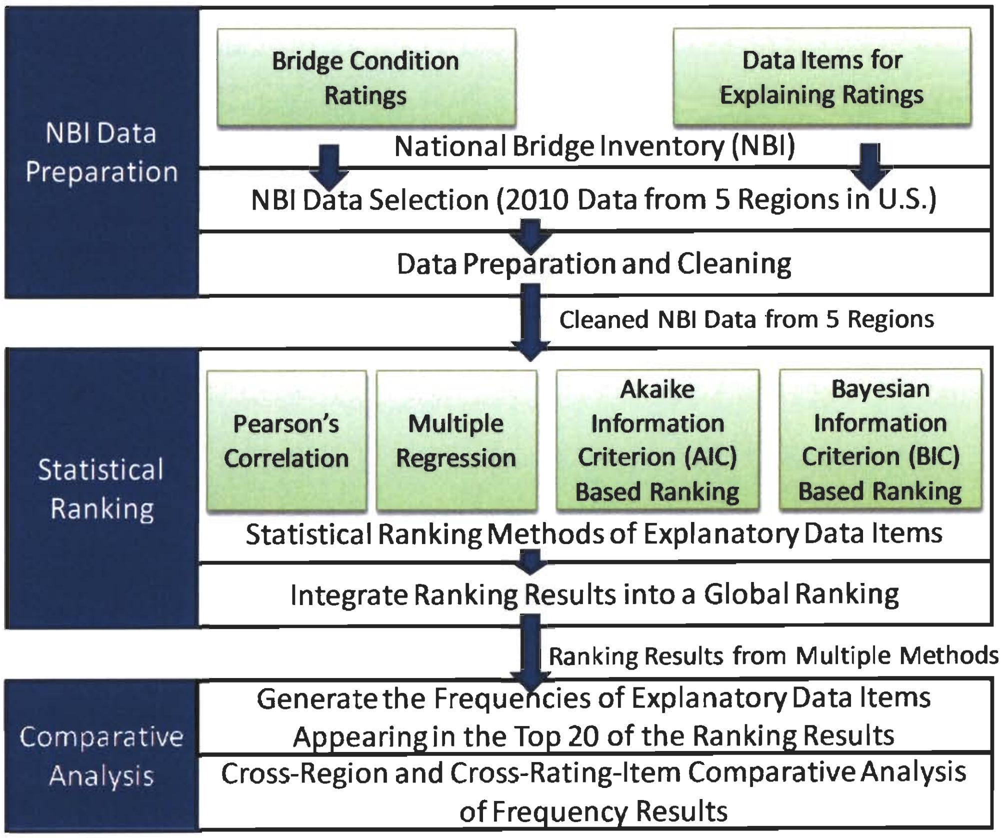

Figure 1: Overview of Comparative Analysis Related to Bridge Condition Evaluation

Stage 2 focuses on identifying which explanatory data items have stronger linear correlations with the rating data items using the four statistical methods: Pearson's Correlation analysis between individual explanatory data items and rating data items; Multiple Regression analysis between multiple explanatory data items and one rating data item; Akaike Information Criterion (AIC) (Akaike 1974) based ranking, and Bayesian Information Criterion (BIC) (Schwarz 1978) based ranking. Subsection "Statistical Methods" below presents details of stage 2. With multiple ranking results generated by multiple methods, a method to integrate multiple results into a global analysis framework was explored by counting how many times each explanatory item appears in the lists generated for each region and rating item to get ranking frequencies of explanatory items and conduct comparative analyses of such frequencies. Section 2.3 "Frequency and Comparative Analysis" presents details of stage 3.

NBI Data Used in this Research

In this research, two aspects were considered while acquiring and preparing NBI data for the frequency analysis of multiple explanatory data items. First, it was necessary to determine which geographic regions to be studied. Second, it was necessary to determine which NBI data items are rating items, and which NBI data items are explanatory items.

According to the coding guide of NBI (Weseman 1995), FHWA divided the U.S. into ten regions and each region has multiple states geographically close to each other. To maximize the environmental diversity of the regions studied in this research, the five regions covering substantial parts of U.S. were selected: Region 1 (Northeast: CT, MA, ME, NH, RI, and VT), Region 4 (Southeast: AL, FL, GA, KY, MS, NC, SC, and TN), Region 5 (Middle-West: IL, IN, MI, MN, OH, and WI), Region 9 (Southwest: AZ, CA, HI, and NV), and Region 10 (Northwest: AK, ID, OR, and WA). This research used regional 2010 NBI data sets released by the Research and Innovative Technology Administration (RITA) of U.S. DOT as parts of "National Transportation Atlas Database" (NTAD hereafter) (www.bts.gov).

In order to classify the data items recorded in NTAD into rating and explanatory data items, all data items documented for each bridge in NTAD were examined. For each bridge, in addition to 116 data items defined in the coding guide of NBI (Weseman 1995), NTAD data include several additional data items (e.g., a binary attribute "Federal" indicating whether the maintenance of a bridge is supported by federal funding). Table 1 shows a classification of all data items in NTAD. Due to the space limits of this paper, this table only shows the labels of these data items, detailed definitions of these items can be found in Appendix A (Weseman 1995).

Table 1 shows that the rating items include four items documenting the conditions of the bridge deck, superstructure, substructure, and channel (including channel protections) using a rating scale from 0 to 9, with 0 indicating the failure, and 9 indicating the perfect condition. The remaining data items in NTAD should be able to serve as valid explanatory items for statistical analysis. Some items, however, was not included into the statistical analysis due to three considerations: 1) some items have missing or non-valid values (e.g. "N" representing "Not Applicable") for more than 90% of the studied bridges in a given region, making the incorporation of them in the statistical analyses non-informative or misleading; 2) some items contain redundant information already covered in other items based on the domain knowledge (e.g.. NBI items 16 and 17 contain the redundant information already in "LONGDD" and "LATDD"); 3) some items are identified as not likely to have substantial correlations with rating items based on the domain knowledge of the author, hence was not included in this study, and will be explored in the future. After eliminating these three categories of data items, the remaining 79 data items in NTAD are considered as valid candidate explanatory items in the frequency analysis. Compared with previous studies, this research included much larger number of candidate explanatory items, hence is a more comprehensive and systematical investigation.

Table 1: Classification of Data Items Recorded in NTAD Files

Rating Data Items	**Items Removed from Data**		
	Non-valid records in < 90% bridges	**Items Identified as Redundant**	**Items Identified as Non-Informative**
Item 58 “Deck”, Item 59 “Superstructure”, Item 60 “Substructure”, Item 61 “Channel and Channel Protection”	STFIPS, REGION, WO, DT, WO_2, SR1SR2, STATUS, DATE_, STPOSTAL, VERSION	ITEM6B , ITEM8 , ITEM13A , ITEM13B , ITEM16 , ITEM75A-B , ITEM76 , ITEM93A-C , ITEM94 , ITEM95 , ITEM97 , ITEM99 , ITEM116 , FUNDED, ITEM67 , ITEM68, ITEM69 , ITEM70 , ITEM71 , ITEM72	ITEM5A-E, ITEM6A, ITEM7, ITEM8, ITEM9, ITEM10, ITEM90, ITEM98A, ITEM98B, ITEM115
Explanatory Data Items			
ITEM2, ITEM3, ITEM10, ITEM11, ITEM12, ITEM19, ITEM20, ITEM21, ITEM22, ITEM26, ITEM27, ITEM28A, ITEM28B, ITEM29, ITEM30, ITEM31, ITEM32, ITEM33, ITEM34, ITEM35, ITEM36A, ITEM36B, ITEM36C, ITEM36D, ITEM37, ITEM38, ITEM39, ITEM40, ITEM41, ITEM42A, ITEM42B, ITEM43A, ITEM43B, ITEM44A, ITEM44B, ITEM45, ITEM46, ITEM47, ITEM48, ITEM49, ITEM50A, ITEM50B, ITEM51, ITEM52, ITEM53, ITEM54A, ITEM54B, ITEM55A, ITEM55B, ITEM56, ITEM63, ITEM64, ITEM65, ITEM66, ITEM91, ITEM92A, ITEM92B, ITEM92C, ITEM96, ITEM100, ITEM101, ITEM102, ITEM103, ITEM104, ITEM105, ITEM106, ITEM107, ITEM108A, ITEM108B, ITEM108C, ITEM109, ITEM110, ITEM111, ITEM112, ITEM113, ITEM114, FEDERAL, LATDD, LONGDD			

Data Cleaning

After determining the FHWA regions to study and identifying the rating and explanatory items, 79 selected explanatory items were further examined. While these remaining 79 data items have valid and informative values for large number of bridges (> 5000 for each region) in NBI, inevitably, there are still some non-valid values of them for some bridges. This research addresses such cases through removing the bridges with missing and non-valid values for any of these 79 data items, known as "data cleaning". Table 2 shows an overview of the data cleaning results (sample R code found in Appendix C). It shows the total numbers of bridges in the original NTAD for all five regions studied in this research, and the number of bridges having complete values (no missing or non-valid values) for all 79 explaining data items. This table also shows that before data cleaning, the original data set for each bridge has 122 attributes.

Table 2: Overview of Data Cleaning Results of NTAD Data Sets for Five Studied Regions

	Original Data		Cleaned Data	
Region	Number of Bridges	Variables	Number of Bridges	Variables
1	22,157	122	14,154	79
4	142,263	122	67,603	79
5	131,087	122	47,403	79
9	47,174	122	8,010	79
10	28,184	122	18,093	79

Frequency and Comparative Analysis

Once the statistical-significance-based ranking results were generated through each of the four statistical methods, there was a need to define a basis by which to compare them. The Correlation and Multiple Regression approaches each generates a ranking list of 79, and 78 items respectively for each studied region, while AIC and BIC methods generate results ranging between 31-61 and 10-38 items respectively, depending on when the forward selection procedure stops(P-value's for each of the explanatory items are found in Appendix C). As these methods have different numbers of ranked explanatory items, this research adopted a frequency analysis method as a baseline to compare how common the candidate explanatory data items are. Specifically, the ranking results are generated from four statistical methods, for five regions, and for four rating items (5 regions × 4 algorithms × 4 rating items = 80 rankings). Table 3 summarizes the top 20 items with the highest appearances.

This method identifies data items based on their frequency in the four algorithms (larger frequencies indicate higher importance). Applying frequency analysis creates a framework which allows for the comparison of the standing of each item regardless of the statistical methods used and the geographic regions.

Table 3: Overall Frequency of Global Top 20 (5 Regions, 4 Rating Items, and 4 Algorithms)

Rank	ITEM	Frequency	Description
1	ITEM27	80	Year Built
1	ITEM91	80	Designated Inspection Frequency
3	ITEM2	77	Highway Agency District
4	ITEM41	76	Structure Open, Posted or Closed to Traffic
4	ITEM43A	76	Kind of material, Main / Approach
6	ITEM106	74	Year Reconstructed
7	ITEM113	72	Scour Critical Bridges
7	ITEM31	72	Design Load
7	ITEM45	72	Number of Spans in Main Unit
10	ITEM108A	70	Type of Wearing Surface
11	ITEM36A	69	Traffic Safety Features - Bridge railings
11	ITEM36D	69	Traffic Safety Features - Approach guardrail ends
11	ITEM48	69	Length of Maximum Span
14	ITEM108C	68	Deck Protection
14	ITEM3	68	County (Parish) Code
14	ITEM37	68	Historical Significance
14	ITEM92B	68	Critical Feature Inspection
18	ITEM107	66	Deck Structure Type
18	ITEM55B	66	Minimum Lateral Underclearance on Right
20	ITEM43B	64	Kind of material, Main / Approach

To study how the relative importance of various explanatory items vary across regions and rating items, a systematic comparative analysis was conducted by grouping frequencies by: a) geographic regions (regional analysis), and b) rating items (rating item analysis). The next section first shows a comparative regional analysis focusing on the differences among the frequencies found in different methods from different regions' data. When conducting regional analysis a maximum frequency of 16 is possible, indicating that a data item was identified as significant by all four algorithms for each of the four rating items. After the regional comparative analysis, a rating item comparative analysis to compare the frequencies of items based on different bridge rating items using different regions' data is presented. In this case, a frequency of 20 is possible indicating that a data item was identified as significant by the four algorithms in each of the five geographic regions.

Frequency Results and Discussions

This section first presents results on analyzing the differences across multiple geographic regions while combining three rating items (listed in Table 1: item 58, Deck; item 59, Superstructure; item 60, Substructure). Figure 2 shows histogram charts synthesizing the frequency comparison results of the top 20 globally frequent items listed in Table 3. Each histogram contains 20 bars representing the frequency of these 20 items in each of the five regions: the minimum value for each is 9 (the least frequent item in Region 9), and the maximum is 16 (upper limit).

Figure 2 shows substantial differences across different geographical regions (aggregating results from four statistical methods and four rating items). The frequency bars generated are dissimilar from one region to another. For all five regions, the charts generally agree that item 27 (Year Built), item 43A (Type of Design, Main / Approach), item 91 (Designated Inspection Frequency), and item 113 (Scour Critical Bridges) are relatively more important; all with a frequency equal to or higher than 14 (85th percentile). Among these four items, items 27, 43A were identified as important in a study conducted in North Dakota using data of 5,289 bridges (Kim and Yoon 2010). The value of items 91 is determined by inspectors based on their visual inspection results. It is therefore intuitive that this item has relatively high correlation with the rating items. The identification of item 113 as significant is also intuitive since scour is an important aspect in the safety evaluation of a bridge's substructure.

A closer examination of such multi-region variations shows that the bars generated have relatively larger variations in Regions 9 and 10, compared with those in Regions 1, 4, and 5. If we only consider Regions 1, 4, and 5, the number of items with frequencies larger than 14 is 12, adding the following items on top of the 4 relatively important items in the previous paragraph: items 31 (Design Load), 36A (Bridge Railings), 41 (Structure Open, Posted or Closed to Traffic), 45 (Number of Spans in Main Unit), 48 (Length of Maximum Span), 106 (Year Reconstructed), 108A (Type of Wearing Surface), and 108C (Deck Protection). Similar to item 91, the value of item 41 is derived by inspectors based on their visual inspection results and item 106 is a secondary measure of the bridges' age. It is therefore intuitive that items 41 and 106 also have relatively high correlations with the rating items in regions 1, 4 and 5. Items 31, 36A, 45, 48 have not been studied in-depth to the best knowledge of the author; this might highlight the potential of three future research directions: 1) explore how the practice of different agencies can influence the bridge deterioration; 2) explore the relationship between bridge size and bridge deterioration; 3) explore the correlations between traffic safety and infrastructure management for better life-cycle management of bridges through improved traffic-safety-oriented design.

Furthermore, the similarity in frequencies observed in regions 1, 4, and 5 contrasted with regions 9, and 10 may indicate that undocumented environmental conditions, such as precipitation and/or humidity, have significant impacts on bridge deterioration. A prominent difference between region 9, and to a lesser extent region 10, and the others is the inferior amount of annual precipitation and humidity for most parts of these regions. This observation indicates that more research might be needed to incorporate weather data as an explanatory item in statistical models.

The frequencies across the four rating items were then analyzed while combining geographic regions and statistical methods. Figure 3 shows histogram charts similar to the previous; each containing the same 20 selected items: the minimum value for each is 13 (least frequent item in item 61"Channel and Channel Protection") and the maximum is 20 (upper limit).

Figure 3 indicates that for all four rating items, all methods generally agree that item 2 (Highway Agency District), item 27 (Year Built), item 41 (Structure Open, Posted or Closed to Traffic), item 91 (Designated Inspection Frequency), item 106 (Year Reconstructed) are relatively more important; all with a frequency equal to or higher than 17 (85th percentile). It must be noted that these items are not describing any physical characteristics of the bridge. However, upon excluding rating item 61 from this frequency analysis, 7 physical items reach a frequency of 17 (or higher) for items 58, 59, and 60 (items 31, 36A/D, 43A, 45, 107, and 108A). Six of these items, excluding item 107 (Deck Structure Type), were found to be significant and were discussed in the regional analysis section of this research. The relatively high frequency of item 107 in rating items 58, 59, and 60 (18, 18, and 17 respectively) suggests that the deck material influences the ratings of both the superstructure and substructure in conjunction with the deck. This research does not conclude as to why this item is related to items 59 and 60 but highlights the opportunity for future research along this direction. The relative similarities in influential items seen in rating items 58, 59, and 60 are also confirmed in Table 4 and Table 5 which show the correlation coefficients within the four rating items and the significance values R^2, also known as coefficient of determination (Efroymson 1960) of all generated statistical analysis results for the rating items. Generally, with the exception of Region 9, correlation coefficients for items 58, 59, and 60 are all larger than 0.4, implying positive correlation between the deck, superstructure, and substructure ratings. Additionally, the resulting coefficients of determination (R^2), Table 5, for

Regions 1, 4, 5, and 10 are all larger than 0.4 implying acceptable model fitting of the rating items (similar significance as reported in Kim and Yoon, 2010) considering that the rating items values are based on visual observations rather than measurable quantities.

It must also be noted that the slight decrease in the coefficient of determination for models generated by AIC and BIC from those generated by multiple regression, Table 5, does not infer worse fitting models; as the criteria for selecting best-fit models in AIC and BIC is not based strictly based on R^2, as discussed previously, but also avoid unnecessary model complexity. That is, taking item 60 in Region 4 as an example, the coefficient of determination values and number of parameters used for models created by multiple regression, AIC and BIC are 0.5365, 0.5361, and 0.5320 and 79, 60 and 38 respectively. The reduction in the coefficient of determination values in AIC and BIC may be due to selecting less parameters, as AIC and BIC both evaluate the value of adding another parameter to the model against the penalty the model assesses when adding another parameter.

Finally, as the frequencies shown in both regional and rating item analysis align well with most conclusions obtained in previous studies, the presence of a few items stand out to be slightly different from the conclusions in previous studies. First, some previous studies highlight traffic data items 29 (Average Daily Traffic, ADT), 109 (Average Daily Truck Traffic), 114 (Future Average Daily Traffic) as important items and studied them in detail (Chase et al 1999; Dunker and Rabbat

1995; Kim and Yoon 2010). In the frequencies generated in this research, none of these items were among the top 20 items selected. These anti-intuitive results need further investigations and maybe field studies in the future to get more detailed traffic data in terms of both time scales and types of vehicles. Another observation in is the absence of longitude and latitude (LONGDD and LATDD). This could be due to the narrow range of the geographic coordinates studied: since each region is analyzed separately, and the geographic extension of each region is limited. In the next step of this research, frequency analysis using the data of the entire U.S. may yield somewhat different results.

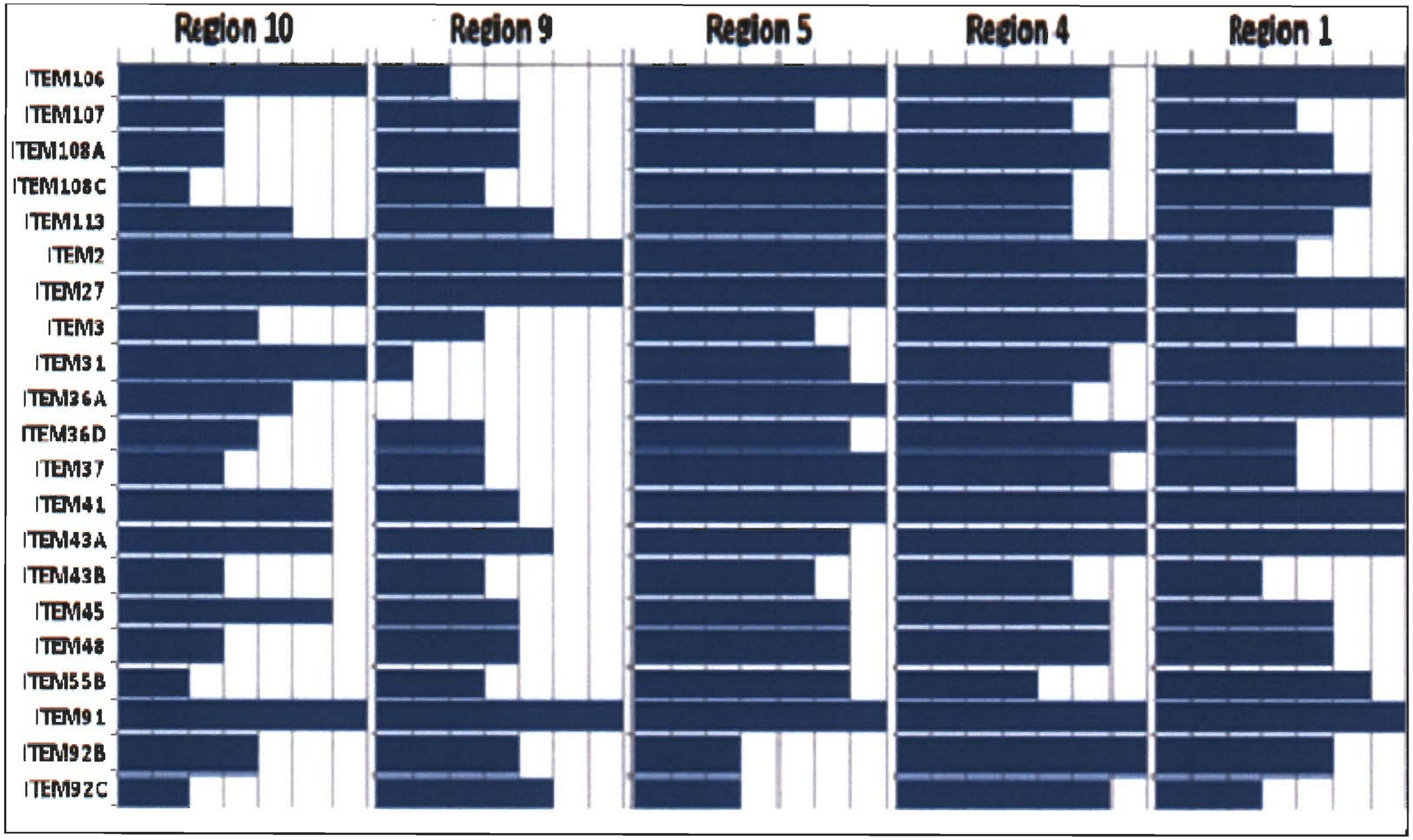

Figure 2: Comparative Analysis of Frequency Result across Geographic Regions

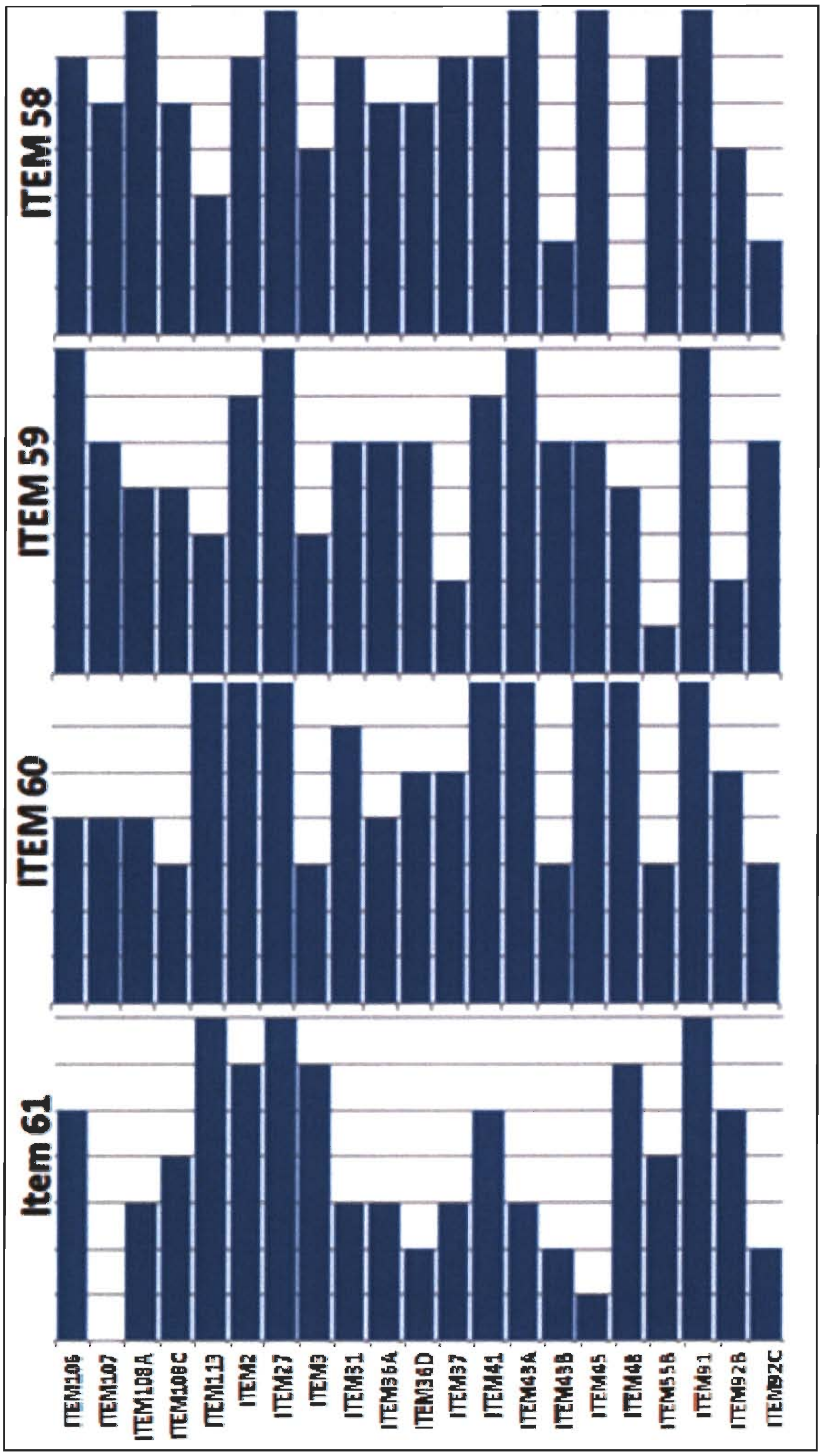

Figure 3: Comparative Analysis of Frequency Result across Rating Items

Table 4: Rating Item Correlations by Region

Region 1	58	59	60	61
ITEM58	1			
ITEM59	0.659	1		
ITEM60	0.543	0.594	1	
ITEM61	0.179	0.209	0.351	1

Region 9	58	59	60	61
ITEM58	1			
ITEM59	0.261	1		
ITEM60	0.177	0.389	1	
ITEM61	-0.008	0.223	0.268	1

Region 4	58	59	60	61
ITEM58	1			
ITEM59	0.794	1		
ITEM60	0.652	0.704	1	
ITEM61	0.426	0.431	0.483	1

Region 10	58	59	60	61
ITEM58	1			
ITEM59	0.659	1		
ITEM60	0.486	0.551	1	
ITEM61	0.176	0.212	0.370	1

Region 5	58	59	60	61
ITEM58	1			
ITEM59	0.830	1		
ITEM60	0.702	0.744	1	
ITEM61	0.469	0.477	0.558	1

Table 5: R-Squared by Region and Rating Item for Multiple Regression, AIC, and BIC

		Region 1	**Region 4**	**Region 5**	**Region 9**	**Region10**
Item 58	Regression	0.4643	0.4241	0.4749	0.3647	0.4205
	AIC	0.4615	0.4238	0.4744	0.359	0.4184
	BIC	0.4282	0.4152	0.4676	0.275	0.3826
Item 59	Regression	0.5168	0.491	0.5526	0.3965	0.4592
	AIC	0.5126	0.4907	0.5524	0.3838	0.4572
	BIC	0.4847	0.4840	0.5362	0.3252	0.4237
Item 60	Regression	0.4581	0.5365	0.5564	0.3476	0.4892
	AIC	0.4554	0.5361	0.5558	0.3337	0.4866
	BIC	0.4231	0.5320	0.5382	0.2797	0.4518
Item 61	Regression	0.2832	0.2545	0.3790	0.3944	0.3086
	AIC	0.2757	0.2538	0.3769	0.3862	0.3027
	BIC	0.2365	0.2495	0.3614	0.3182	0.2826

Conclusion and Future Research

In this chapter, four statistical methods helped to identify the cross-region variations of the relative importance of 79 candidate explanatory data items for explaining the values of four bridge condition rating items in NBI. Comparative analysis across regions and bridge condition rating items shows that: 1) all methods agree that items 27(Year Built) and 91 (Inspection Frequency) have relatively higher correlations with rating items and hence are relatively more important for bridge condition rating; 2) Among the five studied regions, Southwestern region 9 (AZ, CA, HI, and NV), and to a lesser extent Northwestern region 10 (AK, ID, OR, and WA), show substantial differences from others in the identification of significant items, showing the potential for further investigations; 3) this research identified that traffic safety features of bridges (NBI Item 36) and highway agency district (NBI Item 2) as important for bridge condition ratings, while these two items were not studied in-depth in previous studies; 4) traffic volume related data items in NBI are not among top 20 most frequent items, highlighting the importance of a better understanding about how the traffic influences bridge deteriorations along with other factors (e.g., funding sources). These observations may guide engineers to conduct more effective and customized data collection for bridge management across the U.S.

In the future, research to further investigate along the following directions based on these findings may be: 1) further explore how bridge deck conditions are influenced by various environmental conditions in region 9 and 10; 2) conduct field investigations for several factors and regions identified as abnormal in this research (e.g., influence of item 107 on substructure); 3) analyze NBI data at the state level and consider re-regioning based on the similarity of influential factors; 4) develop customized strategies of data collection and bridge management for different regions in U.S. based on the findings of this research.

CHAPTER V

CASE II - PRIORITIZING INFLUENTIAL FACTORS ON LASER SCANNED DATA QUALITY

As previously mentioned, there are numerous factors that influence the quality of a point cloud. Some may be internal; defined by the laser scanner capabilities, such as angular resolution, or external; defined by environment and physical parameters, such as scanning distance and object color. Understanding the extent of their impacts on data qualities will allow engineers to better plan future scans through prioritizing data collection parameters for a given data set. That is, once the scan environment is defined engineers are able to configure the scan parameters to maximize the data quality and minimize the scanning time (by limiting the number of scans. In this chapter, a statistical approach is used to further understand how various parameters influence scan quality.

Literature Review

There have been many studies exploring the possible applications and limitations of laser scanners. Error is typically used as a measure of a scanners limitation based on the dissimilarity between the scanned object and the point cloud. Several studies have explored scanned data quality and scanning improvements as they relate to object based errors (Boehler et al, 2003, Ingensand, 2006), scan-time reduction (B. Akinci et al. 2006; E. Latimer et al. 2004) and scan density (Kizilta et al. 2008).

Object based errors are those directly influenced by the properties of the object scanned. Laser scanners calculate the location of an object based on the signal reflected back from it, therefore, the interaction between the signal and the surface influences the quality of the collected data point. The reflective ability (albedo) directly influences the strength of the returning signal where, typically, white surfaces return a much stronger signal than black surfaces (Boehler et al, 2003, Ingensand, 2006 and Tang et al, 2009). An experiment conducted by Clark and Robson scanned a 24 color checker board at various distances concluding that range errors are highly correlated with surface colors. Similarly, Soudarissanane et al. explored the effects of scanning plywood and medium density fiber boards at different angles. The object's texture/roughness was found to be a significant source of range error.

Conversely, surfaces with exceptionally high reflectivity also cause measurement errors. Voegtle and Wakaluk scanned common construction materials and found that for highly reflective materials (polished marble, granite and metal plates) range data could not be recorded by the scanner due to saturation effects which overwhelm the laser scanner's sensors.

Scan time reduction studies are mainly focused on minimizing the interruptions to construction activities to allow for data collection (B. Akinci et al. 2006; E. Latimer et al. 2004). The reduction in scanning time is created by selecting optimal locations that collect large amounts of data without occlusions. No studies have been conducted relating scan time with optimal scanning conditions and scanner settings.

Scan density has also been identified as a source of influence for data quality. Scan density depends on the scanner properties, i.e. the angular resolution, but also on the scanning geometry, i.e. the incidence angle and the distance of the object from the scanner (Soudarissanane et al. 2007); as the range and incidence angle increase the scan density decreases (Lindenbergh et al., 2005).

These studies all identify certain variables of a scan as influential by measuring the error between the scanned and physical objects and some even identify several of these variables as influencing scan quality. However, no studies have been completed that quantify or rank a variety of physical and environmental variables on scan quality, and systematical error modeling of laser scanners is also limited.

Moreover, systematical error modeling of laser scanners is also limited. Prediction models may advance scan planning and reduce redundant or useless scans by quantifying likely scan errors based on the scanning environment and scanner configuration.

Methodology

Figure 4 shows an overview of the research methodology of this chapter. This exploration has four main stages: 1) Data Collection and Extraction; 2) Data Preparation; 3) Statistical Ranking; and 4) Comparative Analysis.

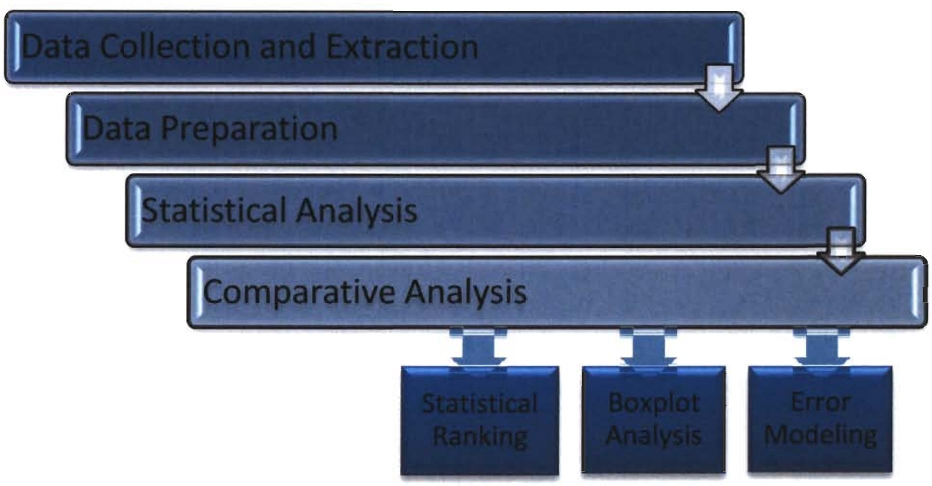

Figure 4: Overview of Laser Scanning Data Analysis Process

In stage 1 cylinder radii and flat surface data were collected using a Leica C10 laser scanner. Cylinders of 3 known sizes where used to create a set of 9 cylinders with variable size and color, Figure 5. For this, each of the three sizes contained black, gray, and white cylinders by applying a thin paper coating. The paper used was of similar roughness and texture to minimize unwanted variability. The shape and color of the objects was selected to capture a wide range of incidence angles with respect to the scanning plane due to the curved shape and a wide range of intensity values; where white returns high intensity while black returns low intensity.

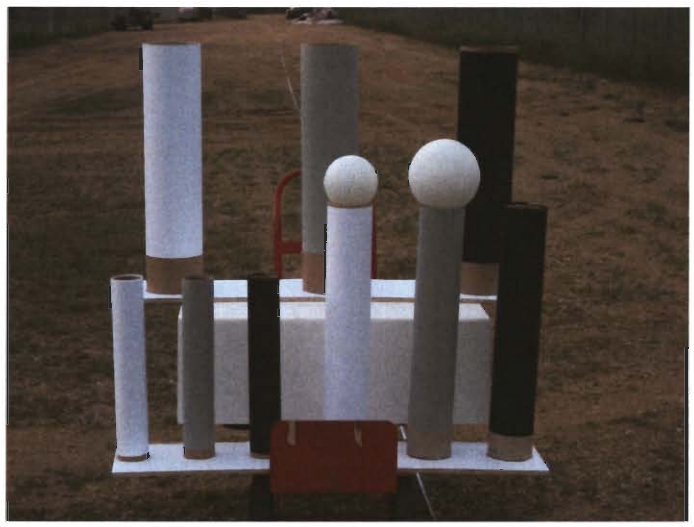

Figure 5: Laser Scanning Data Collection Experimental Setup

Scans where completed at distances of 5, 10, 15, 20, 30, 40, 50, 70, 90, 120 and 150m where most distance contained a low, med, high, and ultra-high resolution scan; scans at 15m and 30m only contained low and medium resolution scans to offset long-range scans that are unusable with these resolution . The same set of scans was completed during both the daytime and nighttime to capture the effects of ambient lighting. Hourly weather data, containing temperature, humidity, and dew point, was also collected from the National Oceanic and Atmospheric Administration (NOAA) through the Kalamazoo International

Airport's facility for the specific time of the experiment and was interpolated to match the times at which the scans were captured.

Once the 80 scans were collected Polyworks® was used to extract the radius of each of the cylinders. This process was repeated 3 times to limit human error while selecting the points to be fitted since the size of the fitted shapes may vary slightly based on the selected points. An examination of the three extractions for day-time scans is summarized in Table 6 which indicates that the average variations between the three extractions are minimal, with an average of .671 mm. The variations where calculated for each cylinder at a given distance and resolution by finding the maximum difference between any of the three extractions.

Table 6: Average Radius Extraction Variations by Size, Color, and Resolution

		Average Variation (mm)
Size	Large	0.502
	Small	0.618
	Medium	0.897
Color	White	0.426
	Gray	0.729
	Black	0.894
Resolution	Low	0.692
	Medium	0.720
	High	0.633
	S. High	0.666
	All	**0.671**

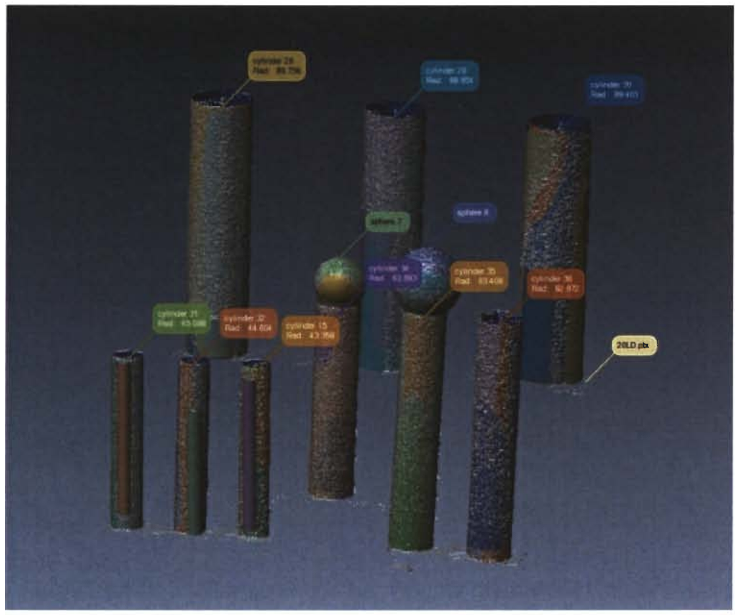

Figure 6: Screenshot of Sample Radius Extraction Using Polyworks

Stage 2 includes substituting non-numeric values with numbers to be used in stage 3; items such as day, night, low resolution, medium resolution etc. were substituted with numbers for the statistical analysis, Table 7. Stage 2 also included substituting error values in cases where the radius was un-extractable. As the scan distances increased the data density decreased leading to un-extractable radii beginning with black cylinders at closer distances, Figure 7 shows the loss of extractable radii as a function of distance. Therefore, if a radius was not extracted the maximum error measured at any distance for that specific cylinder was substituted as the error. That is, if cylinder 5 (C5) at 50m was un-extractable the maximum extracted error for C5 at any other distance was used.

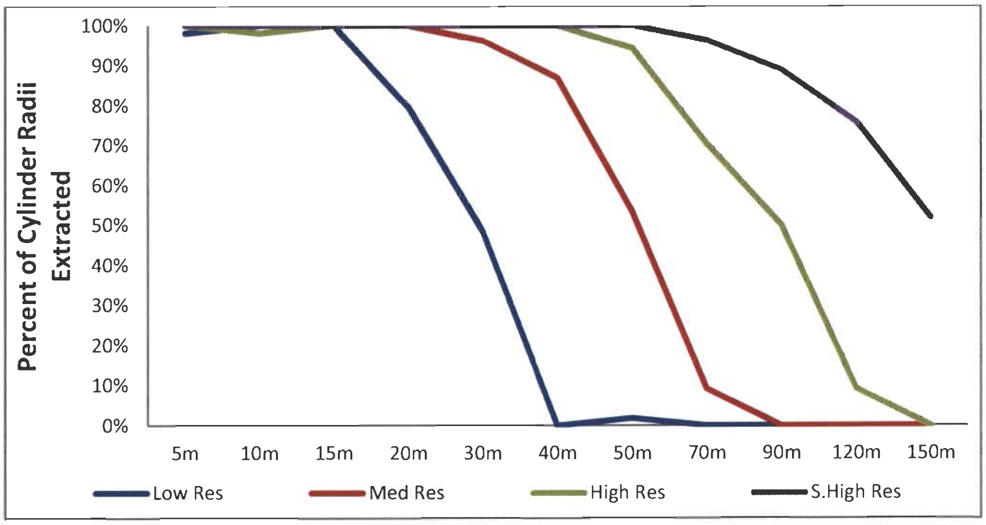

Figure 7: Cylinder Radius Extraction Rates at Various Distances

Table 7: Sample of Data Used for Statistical Analysis

Obj.	Dist (m)	Time	Res.	Color	Measured. Radius (mm)	True Radius (mm)	Temp (F°)	Dew Point	Relative Hum. (%)	Error (mm)
C1	5	1	0	0	89.08	88.52	80	61	52	0.5
C2	5	1	0	1	88.37	88.65	80	61	52	0.28
C3	5	1	0	2	88.09	89.03	80	61	52	0.94
C4	5	1	0	0	43.71	44.25	80	61	52	0.53
C5	5	1	0	1	44.40	44.20	80	61	52	0.20
C6	5	1	0	2	44.15	44.30	80	61	52	0.14
C7	5	1	0	0	94.21	63.28	80	61	52	30.91
C8	5	1	0	1	62.59	63.38	80	61	52	0.78
C9	5	1	0	2	62.08	63.21	80	61	52	1.12
C1	70	0	3	0	89.03	88.52	73	58	59	0.50
C2	70	0	3	1	89.54	88.65	73	58	59	0.88
C3	70	0	3	2	68.96	89.03	73	58	59	20.06
C4	70	0	3	0	44.60	44.25	73	58	59	0.34
C5	70	0	3	1	43.30	44.20	73	58	59	0.87
C6	70	0	3	2	10.08	44.30	73	58	59	34.21
C7	70	0	3	0	63.70	63.28	73	58	59	0.40
C8	70	0	3	1	62.90	63.38	73	58	59	0.49
C9	70	0	3	2	62.26	63.21	73	58	59	0.94

Where:

Time: Day = 1, Night = 2
Resolution: Low = 0, Medium = 1, High = 2, Super High = 3
Color: White = 0, Gray = 1, Black = 2

Stage 3 focuses on identifying the variables that have stronger linear correlations with the measured error, similar to Chapter 2, using four statistical methods: Pearson's Correlation analysis between individual explanatory data items and rating data items; Multiple Regression analysis between multiple explanatory data items and one rating data item; Akaike Information Criterion (AIC) (21) based ranking, and Bayesian Information Criterion (BIC) (22) based ranking.

Finally, in stage 4 further analyses are completed using boxplots to study the relationship between the measured error and various variables culminating in a model to predict error given scan conditions.

Analysis

Statistical Ranking

Similar to the statistical NBI data ranking discussed previously, Pearson's Correlation, Multiple Regression, Akaike Information Criterion (AIC) based ranking, and Bayesian Information Criterion (BIC) based ranking where used to study the correlation between the scan condition variables and the error in radius measurement. A summary of the statistical ranking results are shown in Table 8 and indicate relative agreement between the four methods used with respect to the top 4 items.

Table 8: Summary of Rankings Obtained from Statistical Analysis

Pearson's Correlation		Multiple Regression	
Rank	**Variable**	**Rank**	**Variable**
1	**Intensity**	1	**Distance**
2	**Distance**	2	**Resolution**
3	**Resolution**	3	**Color**
4	**Color**	4	**Intensity**
5	Relative Humidity	5	Dew Point
6	Temperature	6	Temperature
7	Radius	7	Radius
8	Dew Point	8	Relative Humidity
9	Time of Day	9	Time of Day
AIC		**BIC**	
Rank	**Variable**	**Rank**	**Variable**
1	**Intensity**	1	**Intensity**
2	**Resolution**	2	**Resolution**
3	**Distance**	3	**Distance**
4	**Color**	4	**Color**
5	Relative Humidity	9*	Dew Point
6	Dew Point	9*	Relative Humidity
7	Radius	9*	Temperature
9*	Time of Day	9*	Time of Day
9*	Temperature	9*	Radius

*** Variables not appearing in ranking lists are assumed to be ranked last**

Previous research has found that the distance between the scanner and the scanned object is significant to scan quality (Soudarissanane et al. 2007 and Tang et al. 2009). The loss of data quality with an increase in distance may be due to the emitted beam width. That is, as the distance increases, regardless of resolution settings) beam width becomes larger causing laser beam width-induced positional uncertainty (Lichti and Jamtsho 2006). The uncertainty is a result of the process by which the laser scanner measures range; where the position is taken to be the center of the emitted beam and can be problematic at edges and tangent to curved objects like cylinders (Lichti 2004).

The presence of Resolution in the top influencing factors also agrees with previous research. The resolution setting on a scanner refers to its angular resolution which is a measure of the spacing between the emitted beams (Kizilta et al. 2008). This spacing varies at different distances; longer distances amplify the spacing and lower the scan quality. In extracting the radii of the cylinders, large beam spacing decreases the number of points that are reflected from the cylinders surface resulting in a virtual cylinder that does not accurately match the true dimensions of the scanned cylinder. The relationship between the number of reflected beams and radius extraction accuracy also explains the identification of Radius as an influential factor in scan quality.

Box Plot Analysis

Descriptive statistics is a method of quantitatively describing the main characteristics of a data set. Typically, five data parameters are represented in boxplots to visually depict the data as follows: the minimum observation, lower quartile, median, upper quartile, and maximum, Figure 8. In this research boxplots were used to analyze the relationship between distance, resolution, radius and color and the error in radius measurement.

Before analyzing the measured errors with respect to single variables a plot was created to visualize the error with respect to size, resolution, and color. Figure 9 shows the measured average error for each cylinder (color and size specific) at a distance of 20m (the furthest distance with a large number of extractions; 95%, error substitutions were not used). At this distance only the small cylinders had less than perfect extractions with 83, 75, and 96% for C4, C5, and C6 respectively all with low resolution scans

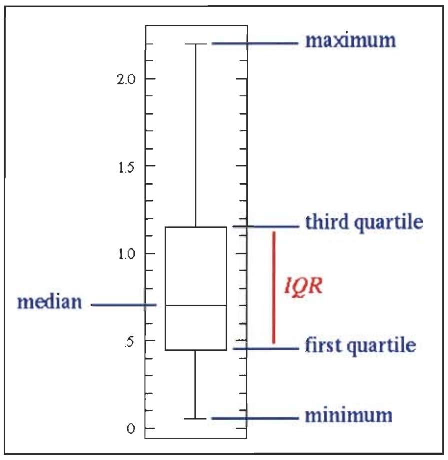

Figure 8: Typical Boxplot Diagram

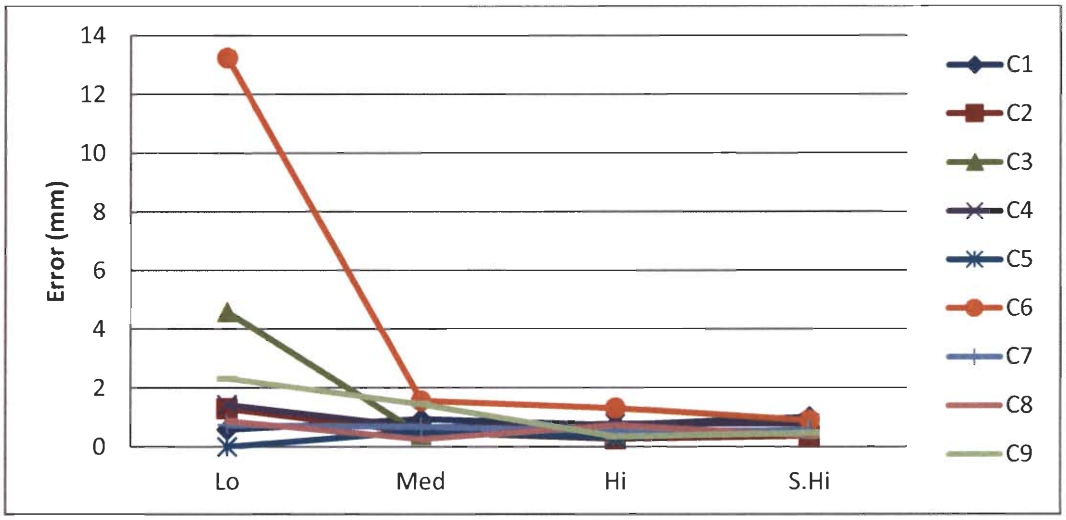

Figure 9: Reduction in Measured Error with Higher Resolutions at 20m

Table 9: Actual Cylinder Dimensions (mm)

ID	Color	Radius (mm)	ID	Color	Radius (mm)
C1	White	88.519	C6	Black	44.297
C2	Gray	88.646	C7	White	63.284
C3	Black	89.027	C8	Gray	63.379
C4	White	44.246	C9	Black	63.207
C5	Gray	44.196			

Since 20m is the distance where low resolution scans begin to create un-extractable data the large error values are expected, however, the importance of color is also observed in the plot considering that C6, C3, and C9 are the black cylinders. The plot also agrees with the statistical rankings by not emphasizing the cylinder size as a source of error; while the three black cylinders have the highest average error they are not in order of size (small, large, then medium).

Boxplots where created to separately analyze the top four variables with respect to measured error. First, the measured errors were compared with the scanning distance to visualize the magnitude of the scanning distance's impact on data quality. For this, boxplots were created to summarize the data at each of the 11 distances. Figure 10 shows the increase in the median error and Q1 as distance increases. The fixed maximum value for error is due to the error substitutions previously mentioned. Also, the sudden increase seen at 30m may be due to only having low and medium resolution scans at that distance which depended heavily on substituted error values. The plot shows a continual increase in all the descriptive statistics but seems to stabilize at 120 and 150m. Both the continual increase and stabilization may be attributed to the increasing number of substituted error as the distance increases.

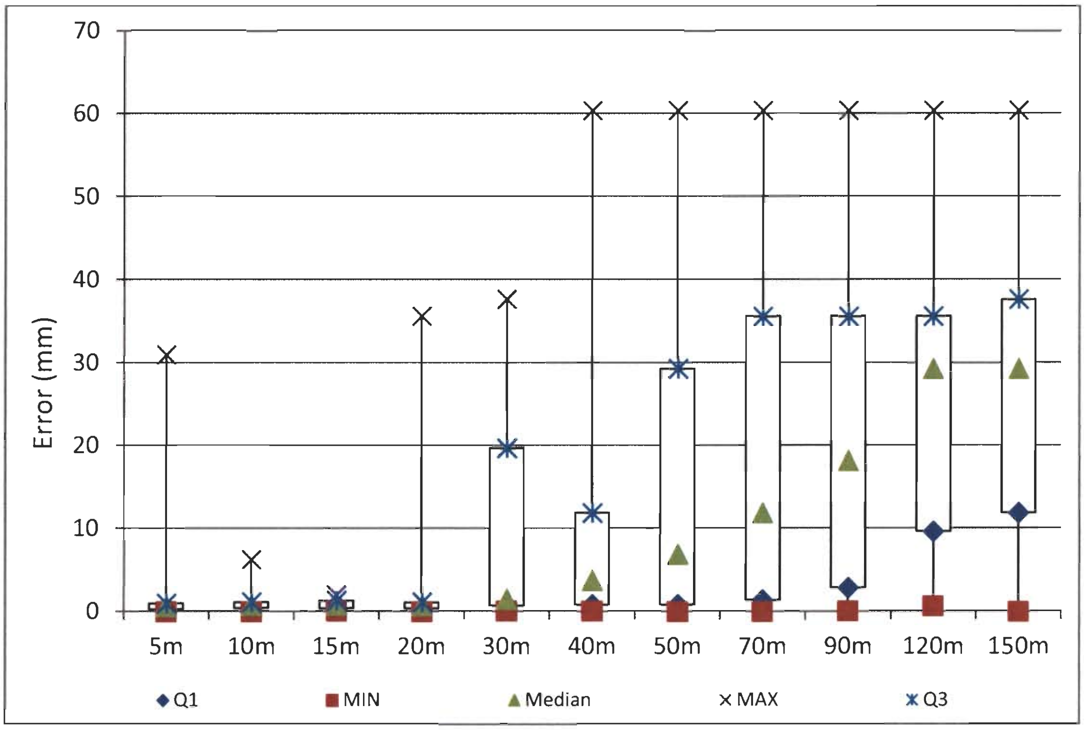

Figure 10: Boxplots at Various Distances with Error Substitutions

The same plot created without substituted error values is seen in Figure 11 and shows a much milder increase in error. However, in this plot the presence of sudden large increases is due to the large errors read at the limits of lower resolutions. The sinusoidal or wavy shape of the boxplots can be explained by the four resolutions used for the scans; as the measured error for each resolution increases with distance until the cylinder radius becomes un-extractable and hence does not contribute to error.

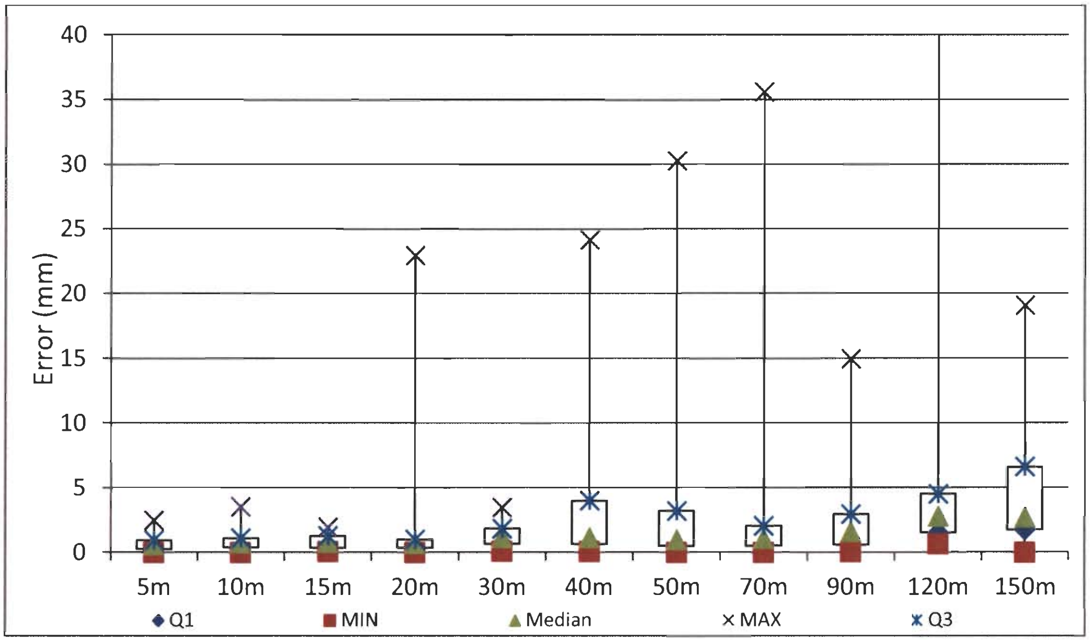

Figure 11: Boxplots at Various Distances without Error Substitutions

Next, Figure 12 A- C shows the results of grouping the scans by resolution, color, and size separately. Figure 12A confirms the importance of resolution on data quality as the median error for super-high resolution scans is considerably lower than that of low resolution (0.05in vs. 0.93in). Figure 12B is also in line with the ranking results as it depicts less of an impact on error from color where the average for white, gray, and black cylinders is 0.053in, 0.054in and 0.37in respectively. The relative differences between the cylinder colors are discussed in more detail in the intensity analysis section.

The relative unimportance of cylinder size in the statistical ranking is also observed in Figure 12C as medium and large cylinders are comparative in their influence to error while small cylinders have marginally larger errors.

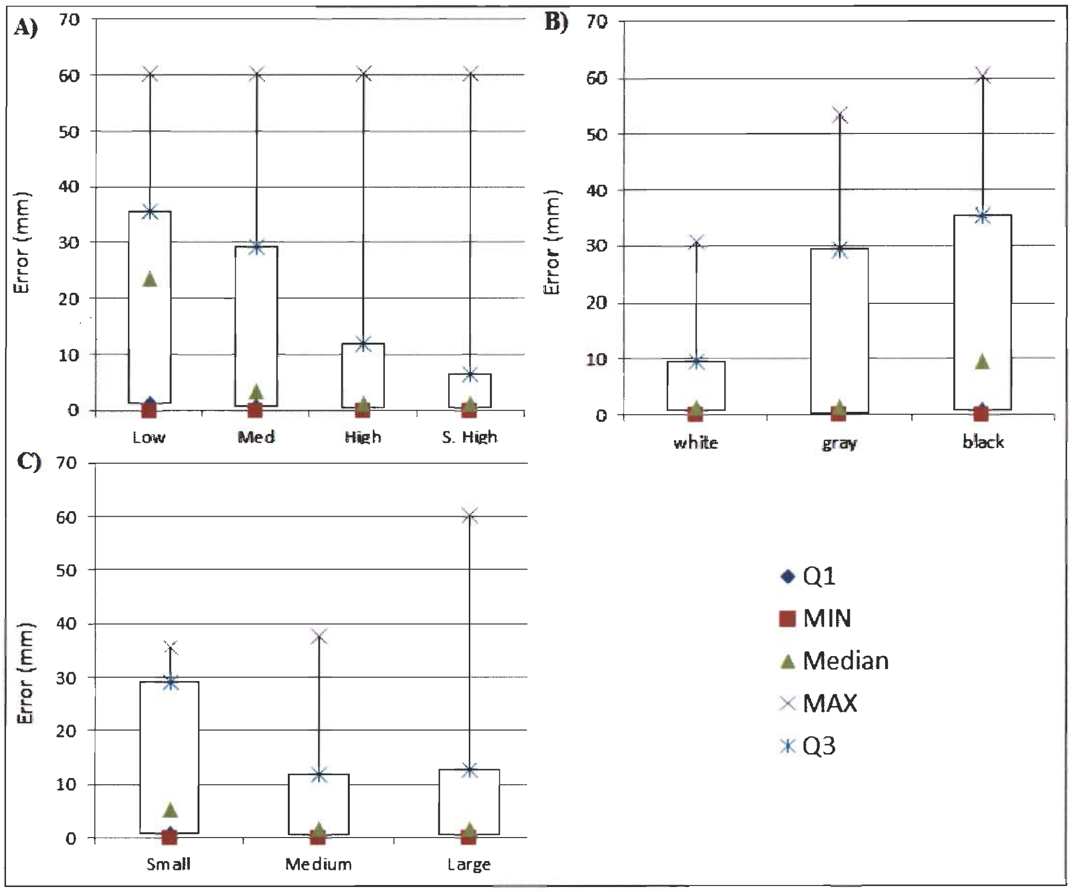

Figure 12: Boxplot Analysis for Scanner Resolution (A), Object Color (B), and Object Size (C)

Plane Fitting

To further explore the influence of object color, scan resolution, and scan distance the flat surface scans were used to analyze deviations from fitted planes created for each scan. A best-fit plane for each of the scans allows for the calculation of the standard deviation of the distance of each point to the plane; where lower standard deviations indicates a better fit. The values obtained from low, medium, high, and super high resolution scans were averaged for each color at each distance since no significant change was observed with different resolutions. As expected, Figure 13 demonstrates the influence of distance on scan quality seen in the increase of the standard deviation with an increase in scan distance. The rate of scan quality deterioration for different colors is also observed as black surfaces' standard deviations increase more rapidly, followed by gray and white surfaces.

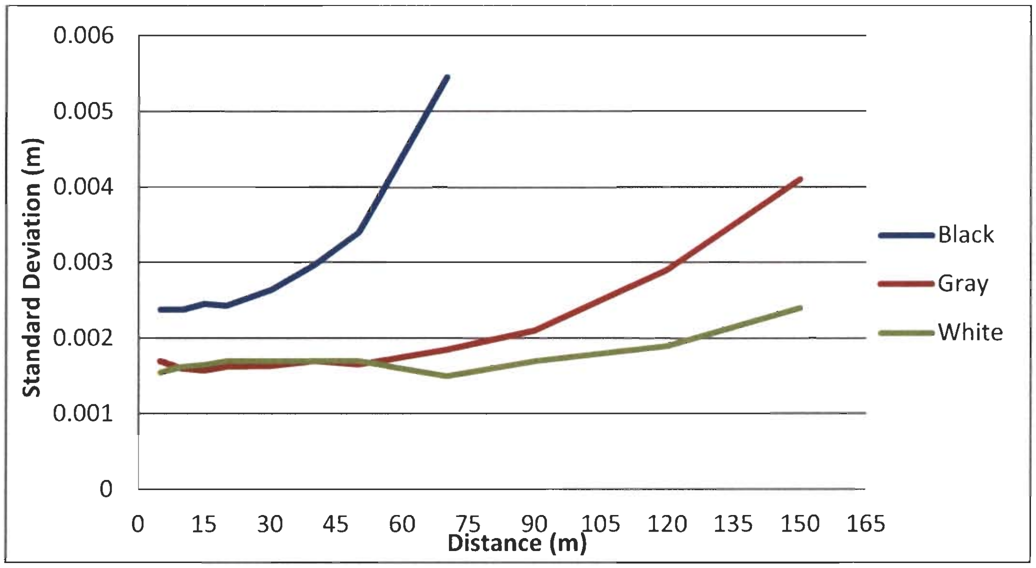

Figure 13: Plane Fitting Standard Deviation

Saturation effects are thought to be the cause of the minor drop in standard deviation for white surfaces at 70m; as saturation effects inflated the values at distances less than 70m where the return laser signal was large.

Intensity Analysis

In order to study the influence of intensity on measured error using boxplots the collected intensity values were normalized and placed into three categories (low, medium, and high intensity). Values in the lower 25% of all observations were considered to have low intensity, the middle 50% (25% to 75%) medium intensity, and the top 25% high intensity; Figure 14.

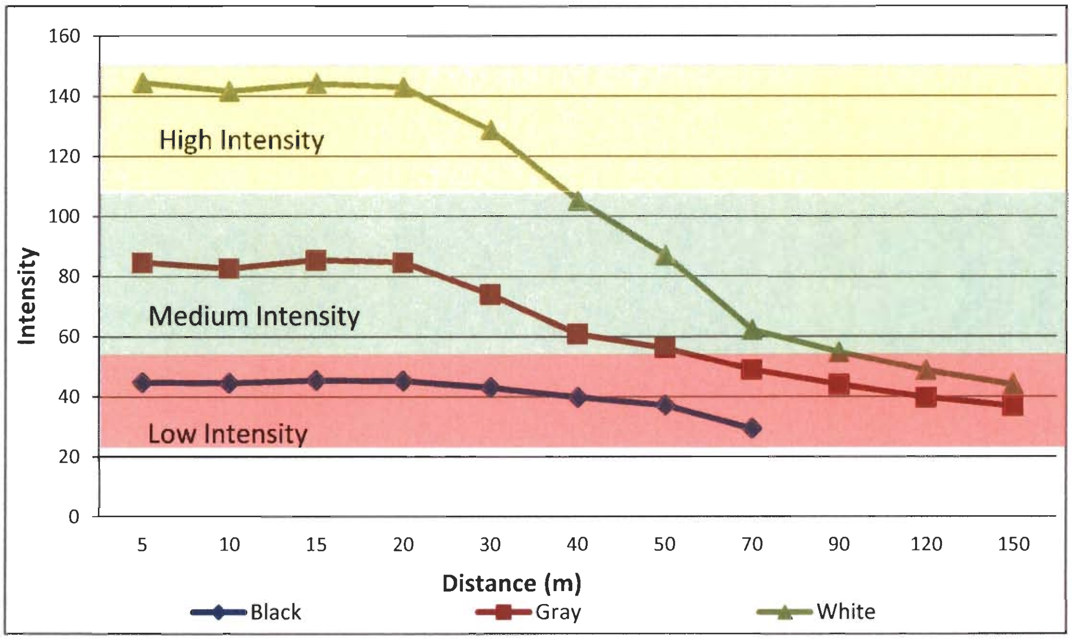

Figure 14: Breakdown of High, Medium, and Low Intensity Ranges

Figure 15 and Table 10 show the boxplot representation of the three intensity classes and the data distribution within each class. The results highlight two aspects of the relationship between intensity and measured error, (1) larger intensity values lead to higher relative accuracy; (2) large intensity values lead to a loss in precision.

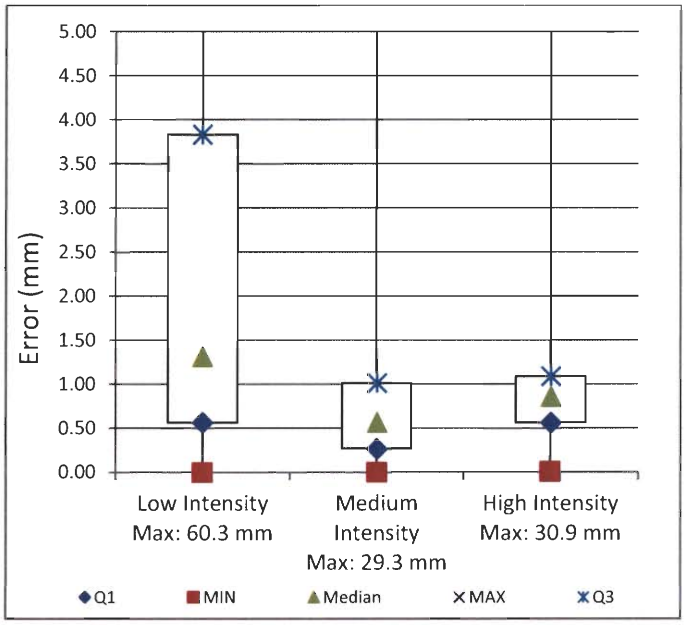

Figure 15: Boxplot Analysis for Scan Intensity

Table 10: Data Distribution for Data Used in Intensity Analysis

	Low Intensity	Medium Intensity	High Intensity
Resolution:			
Low	63	63	63
Medium	90	108	72
High	144	126	54
S. High	252	126	54
Distance:			
5m	72	72	72
10m	72	72	72
15m	18	18	18
20m	72	72	72
30m	9	9	9
40m	54	108	0
50m	72	36	0
70m	72	36	0
90m	36	0	0
120m	36	0	0
150m	36	0	0
Ave Dist	48.9	27.7	12.6
Color:			
White	54	126	243
Gray	126	297	0
Black	369	0	0

The increase in accuracy is seen in the reduction of the interquartile range as intensity increases; an expected result. However, the increase in error seen between medium and high intensity values is unexpected and can only be attributed saturation effects; since all the values in the "high intensity" category were collected for white cylinders within 30m (maximum intensity).

Saturation effects have been seen in high reflectivity surfaces where the internal sensor of a laser scanner is saturated with the signal causing an uncorrected signal recognition and preventing a precise time of flight estimation (Pesci and Teza, 2008; Voegtle and Wakaluk 2009).

Error Modeling

In this section a two-step model to predict the unknown scan parameters given environmental variables is outlined, Figure 16,where step (1) identifies the object's intensity based on color and anticipated scanning distance; and step (2) selects the optimal regression model based on the identified color to predict any single unknown parameter (scan error, scanning resolution, etc.).

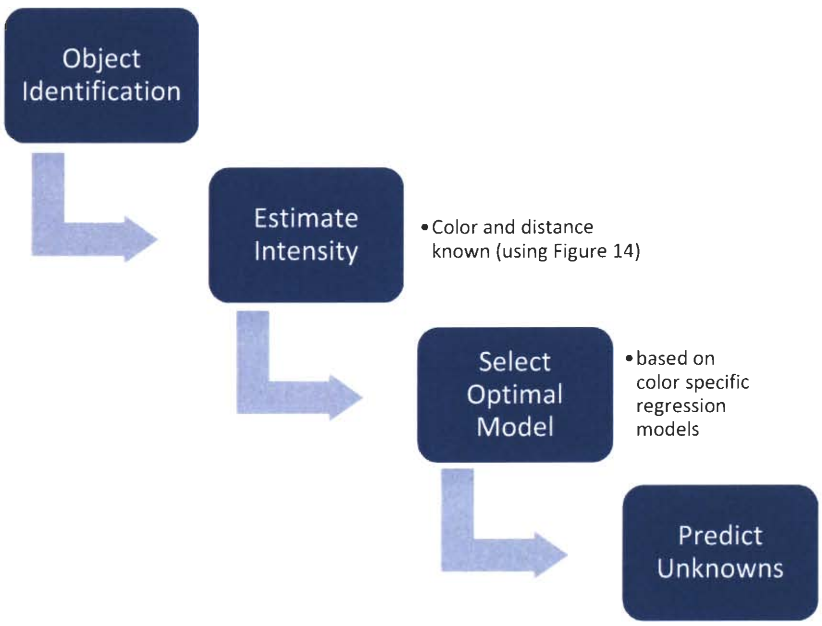

Figure 16: Two-Step Model Flowchart

A two-step model was chosen based on the inaccuracy in using a single prediction model; since the influence of object color varies for the three colors used. Correspondingly, of the nine scan variables studied, object color, dew point, scan distance, object radius, relative humidity, scan resolution, temperature, and time of day can all be identified without the use of a preliminary scan, unlike intensity which is only found using a laser scanner. Therefore, step one is necessary to approximate the object's intensity given scanning distance and object color using Figure 14.

Step two is the use of a prediction model created using a 4-fold model and the four variables identified in Table 8, in which the dataset is randomly partitioned into 4 subsamples; where 3 subsamples are used as training sets using a regression analysis and 1 as a validation set. That is, for each dataset (color) 4 models are created using each subset as a validation set at least once. The sample model summary is shown in Table 11 and is in the form of equation 2. A more detailed summary of the models is found in Appendix D.

$$\text{Error} = \text{Intercept} + \beta_1\,\text{Distance} + \beta_2\,\text{Resolution} + \beta_3\,\text{Color} + \beta_4\,\text{Intensity} \quad (5)$$

Table 11: Prediction Model Coefficients

	Variable	All Colors	White	Gray	Black
	Intercept	0.4410	0.7703	1.1408	0.8595
β_1	Distance	0.0064	0.0003	0.0033	0.0054
β_2	Resolution	-0.1909	-0.1187	-0.2083	-0.0741
β_3	Color	0.1261	0	0	0
β_4	Intensity	-0.0030	-0.0042	-0.0109	-0.0160

Figure 17 shows the boxplots created using the 4-fold prediction models for both the entire dataset and for each color separately; where the Y-axis represents the difference between the measured error and the predicted error using the models.

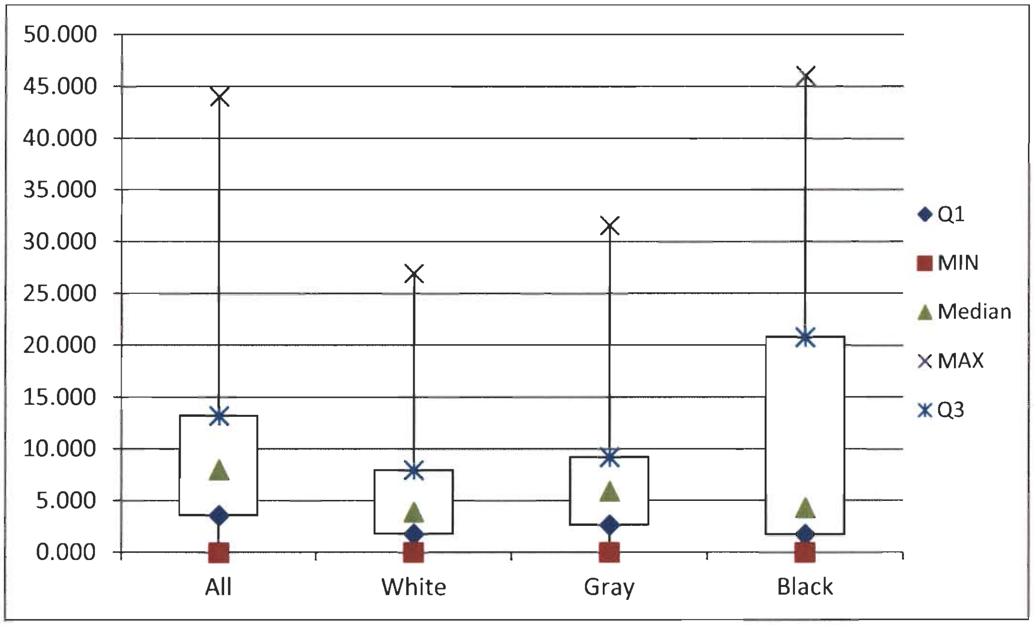

Figure 17: Prediction Model Error Comparison

The above figure establishes that scan error is better predicted for white objects (higher intensity) than for black ones (lower intensity). The relatively large interquartile range for black objects demonstrates the inability to accurately predict scan errors for black objects using linear models. This may be due to either (1) the large number of error substitutions for this subset (black); (2) that the influence of color on scan error is non-linear which is unlikely.

Therefore, a two-step model is recommended in predicting any single scan parameter (scanning distance, scan resolution, or scan error). The model may be used to:

1- Determine scan resolution given a specific scan error tolerance
2- Predict scan error for a given configuration
3- Select the optimum scanning distance using optimization

Conclusion and Future Research

In this study a detailed analysis of variables influencing scan quality was completed. The study utilized four statistical methods (correlation, multiple regression, AIC, and BIC) to rank the variables based on their influence and importance to the quality of a scan and found intensity to be the most influential, followed by scan resolution and distance. The study also confirmed that saturation effects negatively influence scan quality at high intensity conditions. Prediction using linear models was also explored and found to be accurate only for bright objects (high intensity). This may be due to the larger number of samples collected for objects with high intensity; since they are visible at longer distances.

Future research may consider exploring (1) a larger variety of object colors; (2) non-linear prediction models; (3) different laser scanners; (4) various incidence angles.

CHAPTER VI

CONCLUSION

Using statistical methods in a civil engineering context is useful in identifying and quantifying the influence of various factors with in a complex system. A better understanding of the effects of various physical and environmental factors acting within a system allows engineers to better allocate time, funds, and manpower were needed.

In the bridge management case, Case I, this research confirmed that bridge deterioration is not driven by the same factors; where some bridge attributes have more of an impact on bridge deterioration than others in different regions. Hence, customized data collection and maintenance procedures that focus on high ranking deterioration items in a specific region would make bridge management more efficient by eliminating redundant inspections of noncritical bridges. Implementing a region specific inspection and maintenance plan that is specifically designed with regional bridge deterioration considerations is a cost effective solution to the limited funds now allocated to DOT's. Reforming these procedures will not undermine the safety of U.S. bridges but will ensure that critical bridge components and features are better examined. Further research to study influential factor on a state-by-state level may lead to an even deeper understanding of influential factors and may urge policy makers to consider re-grouping states into regions with similar factors influencing their bridges.

In the laser scanning case, Case II, data analysis was used to rank the importance of 9 physical and environmental variables influencing scan quality. While other studies have identified some of these variables as important in scan quality, none have comprehensively studied and ranked such a large group of variables at once. Such a range of variables studied gives engineers a grasp of the impact many site conditions may have on their scans and allows them to better plan around them.

Furthermore, this research created an error prediction model that allows engineers to plan scans based on errors; where engineers can predict the scan error based on scan distance, object color, and scanner configuration. Using this model will ultimately reduce scan times by making scan planning more effective as well as by eliminating situations where scans must be re-captured due to large inaccuracies.

This thesis demonstrates the effectiveness of data analysis for civil engineering applications as shown in the two cases. Whether for historic data, NBI database, or for new and emerging technology, laser scanner, the four statistical methods are able to quantify and rank the influence of various factors to be used to better manage and plan future projects.

REFERENCES

Ahlborn, T. M., et al. (2010). The State-of-the-Practice of Modern Structural Health Monitoring for Bridges: A Comprehensive Review. Publication Task 2 Report of DTOS59-10-H-00001, USDOT Research and Innovative Technology Administration.

Akaike, H. (1974). A New Look at the Statistical Model Identification. Automatic Control, IEEE Transactions on, Vol. 19, No. 6, pp. 716–723.

Becerik-Gerber, B., Jazizadeh, F., Kavulya, G., & Calis, G. (2011). Assessment of target types and layouts in 3D laser scanning for registration accuracy. Automation in Construction, 20(5), 649-658.

Bolukbasi, M., J. Mohammadi, and D. Arditi (2004). Estimating the Future Condition of Highway Bridge Components Using National Bridge Inventory Data. In Practice Periodical on Structural Design and Construction, Vol. 9, No. 1, pp. 16-25.

Burgueño, R., Z. Li. (2008). Identification of Causes and Development of Strategies for Relieving Structural Distress in Bridge Abutments. Publication Research Report (RC-1508) for MDOT under Contract No. 2002-0532. Dept. of Civil and Environmental Engineering, Michigan State University.

Burnham, K. P., and D. R. Anderson (2002). Model Selection and Multimodel Inference : A Practical Information-Theoretic Approach. Springer, New York.

Chase, S. B., C. Nutakor, and E. Small (1999). An In-Depth Analysis of the National Bridge Inventory Database Utilizing Data Mining, GIS and Advanced Statistical Methods. In Eight Transportation Board Conference on Bridge Management, number C-6; I BMC-04, Vol. 7, pp. 1-17.

Chase, S., B. Harris, E. Small, and C. Nutakor (1998). Associating Factors Influencing Deterioration and Load Carrying Capacity With Bridges: A GIS Approach. In Proceedings of ESRI International User Conference 1998. CD-ROM. ESRI, Redlands, CA.

Clark, J., Robson, S. (2004). Accuracy of measurements made with CYRAX 2500 laser scanner against surfaces of known colour. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Commission IV, Part B4, pp. 1031-1037

Dalgaard, P. Introductory Statistics with R (2nd edition). Springer Verlag, New York, 2008.

Dunker, K. F. and B. G. Rabbat (1995). Assessing Infrastructure Deficiencies: the Case of Highway Bridges. In Journal of Infrastructure Systems, Vol. 1, No. 2, pp. 100–119.

Dunker, K. F. and B. G. Rabbat (1995). Assessing Infrastructure Deficiencies: The Case of Highway Bridges. In Journal of Infrastructure Systems, Vol. 1, No. 2, pp. 100-119.

Efroymson, M. A. (1960). Multiple regression analysis. In Mathematical methods for digital computers, Vol. 1, A. Ralston and H. Wilf, Eds. Wiley: New York, pp. 191–203.

Gayawan, E. and R. A. Ipinyomi (2009). A Comparison of Akaike, Schwarz and R Square Criteria for Model Selection Using Some Fertility Models. Australian Journal of Basic Applied Sciences, 3(4)

Glaser, S. D., and A. Tolman (2008). Sense of Sensing: From Data to Informed Decisions for the Built Environment. Journal of Infrastructure Systems, Vol. 14, No. 1, pp. 4-14.

Gokey, J., N. Klein, C. Mackey, J. Santos, A. Pillutla, and S. Tucker (2009). Development of a Prioritization Methodology for Maintaining Virginia's Bridge Infrastructure Systems. In Proceedings of the 2009 IEEE Systems and Information Engineering Design Symposium, University of Virginia, Charlottesville, VA, USA, pp. 252-257.

Karlaftis, M. (2005). A Web-Supported National Bridge Inventory Management Tool. In Public Works Management & Policy, Vol. 9, No. 3, pp. 248-258.

Kim, Y. J. and D. K. Yoon (2010). Identifying Critical Sources of Bridge Deterioration in Cold Regions through the Constructed Bridges in North Dakota. In Journal of Bridge Engineering, Vol. 15, No. 5, pp. 542-552.

Kiziltas, S., Akinci, B., Ergen, E., Tang, P., & Gordon, C. (2008). Technological assessment and process implications of field data capture technologies for construction and facility/infrastructure management. ITcon, 13(April), 134-154

Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. In Sociological Methods & Research, Vol. 33, No. 2, pp. 188-229.

Li, Z. and R. Burgueno(2010). Using Soft Computing to Analyze Inspection Results for Bridge Evaluation and Management. In Journal of Bridge Engineering, Vol. 15, No. 4, pp. 430-438.

Lichtenstein, A. G. (1993). The Silver Bridge Collapse Recounted. Journal of performance of constructed facilities, Vol. 7, No. 4, pp. 249–261.

Lichti, D. D. (2004). A resolution measure for terrestrial laser scanners. Knowledge Creation Diffusion Utilization, 34(Part XXX), 552-558.

Lichti, D. D. (2007). Error modeling, calibration and analysis of an AM–CW terrestrial laser scanner system. Journal of Photogrammetry & Remote Sensing, 307-324.

McClure, S. and K. Daniell (2010). Development of User-Friendly Software Application for Extracting Information from National Bridge Inventory Source Files. In Transportation Research Record: Journal of the Transportation Research Board, Vol. 2202, No. 1, pp. 137–147.

Morcous, G. (2002). Modeling Bridge Deterioration Using Case-Based Reasoning. In Journal of Infrastructure Systems, Vol. 8, No. 3, pp. 86-95.

Pesci, A., & Teza, G. (2008). Terrestrial laser scanner and retro-reflective targets: an experiment for anomalous effects investigation. International Journal of Remote Sensing, 29(19), 5749-5765

RITA, BTS | National Transportation Atlas Database. USDOT, 2011. Available: http://www.bts.gov/publications/national_transportation_atlas_database/. Accessed: 07-Jul-2011.

Schwarz, G. (1978). Estimating the Dimension of a Model. In The Annals of Statistics, Vol. 6, No. 2, pp. 461–464.

Soudarissanane, S., Van Ree, J., Bucksch, A. and Lindenbergh, R. (2007). Error budget of terrestrial laser scanning: influence of the incidence angle on the scan quality. In: Proceedings 3D-NordOst, Berlin, Germany.

Soudarissanane, S., Lindenbergh, R., & Gorte, B. (2008). Reducing the error in terrestrial laser scanning by Optimizing the measurement set-up. ISPRS Congress (Vol. 5, p. 615ff)

Su, Y. Y., Hashash, Y. M. A., & Liu, L. Y. (2006). Integration of Construction As-Built Data via Laser Scanning with Geotechnical Monitoring of Urban Excavation. Journal of Construction Engineering and Management, 132(12), 1234-1241.

Tang, P. and B. Akinci (2009). Extracting Surveying Goals from Point Clouds to Support Construction and Infrastructure Inspection. In Building a Sustainable Future Proceedings of the 2009 Construction Research Congress, pp. 1164-1173.

Tokdemir, O. B., C. Ayvalik, and J. Mohammadi (2000). Prediction of Highway Bridge Performance by Artificial Neural Networks and Genetic Algorithms. In Proceeding of the 17th International Symposium on Automation and Robotics in Construction (ISARC). CD-ROM. Taipei, Taiwan.

Valenzuela, S., H. de Solminihac, and T. Echaveguren (2010). Proposal of an Integrated Index for Prioritization of Bridge Maintenance. In Journal of Bridge Engineering, Vol. 15, No. 3, pp. 337-343.

Wang, X., W. Dou, S.-E. Chen, W. Ribarsky, and R. Chang (2010). An Interactive Visual Analytics System for Bridge Management. In Computer Graphics Forum, Vol. 29, No. 3, pp. 1033-1042.

Weseman, W. A. (1995). Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges. Publication FHWA-PD-96-001. FHWA, U.S. Department of Transportation, D.C.

Appendix A

NBI Data Item Description

NBI Item Description

Item #	Description
1	State Code
2	Highway Agency District
3	County (Parish) Code
4	Place Code
5	Inventory Route
6	Features Intersected
7	Facility Carried by Structure
8	Structure Number
9	Location
10	Inventory Route, Minimum Vertical Clearance
11	Kilometer Point
12	Base Highway Network
13	LRS Inventory Route, Subroute Number
19	Bypass, Detour Length
20	Toll
21	Maintenance Responsibility
22	Owner
26	Functional Classification of Inventory Route
27	Year Built
28	Lanes On and Under the Structure
29	Average Daily Traffic
30	Year of Average Daily Traffic
31	Design Load
32	Approach Roadway Width
33	Bridge Median
34	Skew
35	Structure Flared
36	Traffic Safety Features
37	Historical Significance
38	Navigation Control
39	Navigation Vertical Clearance
40	Navigation Horizontal Clearance
41	Structure Open, Posted or Closed to Traffic
42	Type of Service
43	Structure Type, Main

Item #	Description
44	Structure Type, Approach Spans
45	Number of Spans in Main Unit
46	Number of Approach Spans
47	Inventory Route, Total Horizontal Clearance
48	Length of Maximum Span
49	Structure Length
50	Curb or Sidewalk Widths
51	Bridge Roadway Width, Curb-to-Curb
52	Deck Width, Out-to-Out
53	Minimum Vertical Clearance Over Bridge Roadway
54	Minimum Vertical Underclearance
55	Minimum Lateral Underclearance on Right
56	Minimum Lateral Underclearance on Left
58	Deck Condition Rating
59	Superstructure Condition Ratings
60	Substructure Condition Ratings
61	Channel and Channel Protection
62	Culverts Condition Ratings
63	Method used to Determine Operating Rating
64	Operating Rating
65	Method used to Determine Inventory Rating
66	Inventory Rating
67	Structural Evaluation Appraisal Ratings
68	Deck Geometry Appraisal Ratings
69	Underclearances, Vertical and Horizontal Appraisal Ratings
70	Bridge Posting
71	Waterway Adequacy Appraisal Ratings
72	Approach Roadway Alignment Appraisal Ratings
75	Type of Work
76	Length of Structure Improvement
90	Inspection Date
91	Designated Inspection Frequency
92	Critical Feature Inspection
93	Critical Feature Inspection Date
94	Bridge Improvement Cost
95	Roadway Improvement Cost
96	Total Project Cost

Item #	Description
97	Year of Improvement Cost Estimate
98	Border Bridge
99	Border Bridge Structure Number
100	STRAHNET Highway Designation
101	Parallel Structure Designation
102	Direction of Traffic
103	Temporary Structure Designation
104	Highway System of the Inventory Route
105	Federal Lands Highways
106	Year Reconstructed
107	Deck Structure Type
108	Wearing Surface/Protective System
109	Average Daily Truck Traffic
110	Designated National Network
111	Pier or Abutment Protection [for navigation]
112	NBIS Bridge Length
113	Scour Critical Bridges
114	Future Average Daily Traffic
115	Year of Future Average Daily Traffic
116	Minimum Navigation Vertical Clearance

Appendix B

Sample R Code for Case I

```
data.clean=function(x) {
  xcol=ncol(x)
  xind=NULL
  for ( j in 1:xcol ) {
    xind=c(xind,which(x[,j]=="" | x[,j]=="*"))
  }
  xind=unique(xind)
  return(x[-xind,])
}

f2num=function(x) return(as.numeric(as.character(x)))



########################## Data Cleaning ######################

load("nbi_region4_2010.RData")
ls()
nbi4=nbi_Region4
names(nbi4)
nbi4=nbi4[,-c(1:2,122:124,126:129,132:133)]

nbi4_58=nbi4[nbi4[,66]!=""&nbi4[,66]!="N",c(66,1:65,71:119,121:122)]
nbi4_59=nbi4[nbi4[,67]!=""&nbi4[,67]!="N",c(67,1:65,71:119,121:122)]
nbi4_60=nbi4[nbi4[,68]!=""&nbi4[,68]!="N",c(68,1:65,71:119,121:122)]
nbi4_61=nbi4[nbi4[,69]!=""&nbi4[,69]!="N",c(69,1:65,71:119,121:122)]
nbi4_62=nbi4[nbi4[,70]!=""&nbi4[,70]!="N",c(70,1:65,71:119,121:122)]
nbi4_stat=nbi4[nbi4[,120]!=""&nbi4[,120]!="N",c(120,1:65,71:119,121:122)]

########################## ITEM59 ######################

nbi=nbi4_59

### nbi.na=NULL
### for ( j in 1:ncol(nbi) ) nbi.na=c(nbi.na,sum(nbi[,j]=="" | nbi[,j]=="N"))

tmp.ind=which(nbi[,92]!="")
nbi.tmp=nbi[tmp.ind,92]
nbi[,92]=rep("N",nrow(nbi))
nbi[tmp.ind,92]=as.character(nbi.tmp)

tmp.ind=which(nbi[,93]!="")
nbi.tmp=nbi[tmp.ind,93]
nbi[,93]=rep("N",nrow(nbi))
nbi[tmp.ind,93]=as.character(nbi.tmp)

tmp.ind=which(nbi[,98]!="")
nbi[,98]=rep("F",nrow(nbi))
nbi[tmp.ind,98]="T"

tmp.ind=which(nbi[,108]!="")
```

```
nbi.tmp=nbi[tmp.ind,108]
nbi[,108]=rep("0",nrow(nbi))
nbi[tmp.ind,108]=as.character(nbi.tmp)

nbi=nbi[,-c(12,18:19,20:21,71:76,77:79,85:87,88:89,91,94,113,114)]

nbi=data.clean(nbi)

nbi=nbi[,-c(2:7,10:13,66,72:73,91)]

#### ITEM4??? ITEM108???

nbi$ITEM59=f2num(nbi$ITEM59)
nbi$ITEM10=as.factor(f2num(nbi$ITEM10)>3000)
nbi$ITEM11=f2num(nbi$ITEM11)/1000
nbi$ITEM19=f2num(nbi$ITEM19)
nbi$ITEM27=f2num(nbi$ITEM27)
nbi$ITEM28A=f2num(nbi$ITEM28A)
nbi$ITEM28B=f2num(nbi$ITEM28B)
nbi$ITEM29=f2num(nbi$ITEM29)
nbi$ITEM32=f2num(nbi$ITEM32)/10
nbi$ITEM34=f2num(nbi$ITEM34)
nbi$ITEM39=as.factor(f2num(nbi$ITEM39)>0)
nbi$ITEM40=as.factor(f2num(nbi$ITEM40)>0)
nbi$ITEM45=f2num(nbi$ITEM45)
nbi$ITEM46=as.factor(f2num(nbi$ITEM46)>0)
nbi$ITEM47=f2num(nbi$ITEM47)/10
nbi$ITEM48=f2num(nbi$ITEM48)/10
nbi$ITEM49=f2num(nbi$ITEM49)/10
nbi$ITEM50A=f2num(nbi$ITEM50A)/10
nbi$ITEM50B=f2num(nbi$ITEM50B)/10
nbi$ITEM51=f2num(nbi$ITEM51)/10
nbi$ITEM52=f2num(nbi$ITEM52)/10
nbi$ITEM53=as.factor(f2num(nbi$ITEM53)>3000)
nbi$ITEM54B=as.factor(f2num(nbi$ITEM54B)>0)
nbi$ITEM55B=as.factor(f2num(nbi$ITEM55B)>0)
nbi$ITEM56=as.factor(f2num(nbi$ITEM56)>0)
nbi$ITEM64=f2num(nbi$ITEM64)/10
nbi$ITEM66=f2num(nbi$ITEM66)/10
nbi$ITEM91=f2num(nbi$ITEM91)
nbi$ITEM92A[nbi$ITEM92A=="N00"]="N"
nbi$ITEM92B[nbi$ITEM92B=="N00"]="N"
nbi$ITEM92C[nbi$ITEM92C=="N00"]="N"
nbi$ITEM96=f2num(nbi$ITEM96)
nbi$ITEM103=as.factor(nbi$ITEM103)
nbi$ITEM106=f2num(nbi$ITEM106)
nbi$ITEM109=f2num(nbi$ITEM109)
nbi$ITEM111=as.factor(nbi$ITEM111)
nbi$ITEM114=f2num(nbi$ITEM114)

item.is.f=NULL
```

```
for ( j in 2:ncol(nbi) ) item.is.f=c(item.is.f,is.factor(nbi[,j]))
nbi=nbi[,c(1,1+which(item.is.f==T),1+which(item.is.f==F))]


cor.tmp=matrix(0,nrow(nbi),ncol(nbi))
for (j in 1:ncol(nbi)) cor.tmp[,j]=as.numeric(nbi[,j])
cor.mat=cor(cor.tmp)
max2=function(x) return(-sort(-x)[2])
which.max2=function(x) return(order(-x)[2])
cbind(names(nbi),names(nbi)[apply(cor.mat,2,which.max2)],apply(cor.mat,2,max2))


#####################    Correlation    ######################

cor.order=order(-abs(cor.mat[1,-1]))
t.test.cor=function(r,n) {
  t.test=r*sqrt((n-2)/(1-r^2))
  p.val=2*(1-pt(abs(t.test),n-2))
  return(list(t.test=t.test,p.val=p.val))
}
cbind(names(nbi)[-1],cor.mat[1,-1],t.test.cor(cor.mat[1,-1],nrow(nbi))$p.val)[cor.order,]


##################### Multiple regression ######################

Full=lm(ITEM59~.,data=nbi)
cbind(names(nbi)[-c(1,21)],anova(Full)$"Pr(>F)"[-79])[order(anova(Full)$"Pr(>F)"[-79]),]

### AIC ###
nbi.aic=step(lm(ITEM59~1, data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=2)
m.aic=lm(ITEM59 ~ ITEM27 + ITEM43A + ITEM41 + ITEM43B + ITEM3 + ITEM91 +
ITEM22 + ITEM31 + ITEM2 + ITEM36A + ITEM30 + ITEM44B + ITEM64 + ITEM65 +
ITEM107 + ITEM36C + ITEM42B + ITEM103 + ITEM45 + ITEM106 + ITEM37 + ITEM44A +
ITEM92C + ITEM113 + ITEM20 + ITEM105 + FEDERAL + ITEM21 + ITEM109 + ITEM36D
+ ITEM26 + ITEM92A + ITEM50B + ITEM52 + ITEM108B + ITEM36B + ITEM34 + ITEM51
+ ITEM12 + ITEM92B + ITEM110 + ITEM108C + ITEM53 + ITEM111 + ITEM32 + ITEM33 +
ITEM47, data=nbi)
anova(m.aic,Full)


### BIC ###
nbi.bic=step(lm(ITEM59~1,
data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=log(nrow(nbi)))
m.bic=lm(ITEM59 ~ ITEM27 + ITEM43A + ITEM41 + ITEM91 + ITEM22 + ITEM36A +
    ITEM43B + ITEM2 + ITEM64 + ITEM65 + ITEM31 + ITEM46 + ITEM55B +
    ITEM36C + ITEM103 + ITEM45 + ITEM106 + ITEM105 + FEDERAL +
    ITEM50B + ITEM20 + ITEM52, data=nbi)
anova(m.bic,Full)

data.clean=function(x) {
  xcol=ncol(x)
```

```
  xind=NULL
  for ( j in 1:xcol ) {
    xind=c(xind,which(x[,j]=="" | x[,j]=="*"))
  }
  xind=unique(xind)
  return(x[-xind,])
}

f2num=function(x) return(as.numeric(as.character(x)))




########################## ITEM58 #######################

nbi=nbi4_58

### nbi.na=NULL
### for ( j in 1:ncol(nbi) ) nbi.na=c(nbi.na,sum(nbi[,j]=="" | nbi[,j]=="N"))

tmp.ind=which(nbi[,92]!="")
nbi.tmp=nbi[tmp.ind,92]
nbi[,92]=rep("N",nrow(nbi))
nbi[tmp.ind,92]=as.character(nbi.tmp)

tmp.ind=which(nbi[,93]!="")
nbi.tmp=nbi[tmp.ind,93]
nbi[,93]=rep("N",nrow(nbi))
nbi[tmp.ind,93]=as.character(nbi.tmp)

tmp.ind=which(nbi[,98]!="")
nbi[,98]=rep("F",nrow(nbi))
nbi[tmp.ind,98]="T"

tmp.ind=which(nbi[,108]!="")
nbi.tmp=nbi[tmp.ind,108]
nbi[,108]=rep("0",nrow(nbi))
nbi[tmp.ind,108]=as.character(nbi.tmp)

nbi=nbi[,-c(12,18:19,20:21,71:76,77:79,85:87,88:89,91,94,113,114)]

nbi=data.clean(nbi)

nbi=nbi[,-c(2:7,10:13,66,72:73,91)]

#### ITEM4??? ITEM108???

nbi$ITEM58=f2num(nbi$ITEM58)
nbi$ITEM10=as.factor(f2num(nbi$ITEM10)>3000)
nbi$ITEM11=f2num(nbi$ITEM11)/1000
nbi$ITEM19=f2num(nbi$ITEM19)
```

```
nbi$ITEM27=f2num(nbi$ITEM27)
nbi$ITEM28A=f2num(nbi$ITEM28A)
nbi$ITEM28B=f2num(nbi$ITEM28B)
nbi$ITEM29=f2num(nbi$ITEM29)
nbi$ITEM32=f2num(nbi$ITEM32)/10
nbi$ITEM34=f2num(nbi$ITEM34)
nbi$ITEM39=as.factor(f2num(nbi$ITEM39)>0)
nbi$ITEM40=as.factor(f2num(nbi$ITEM40)>0)
nbi$ITEM45=f2num(nbi$ITEM45)
nbi$ITEM46=as.factor(f2num(nbi$ITEM46)>0)
nbi$ITEM47=f2num(nbi$ITEM47)/10
nbi$ITEM48=f2num(nbi$ITEM48)/10
nbi$ITEM49=f2num(nbi$ITEM49)/10
nbi$ITEM50A=f2num(nbi$ITEM50A)/10
nbi$ITEM50B=f2num(nbi$ITEM50B)/10
nbi$ITEM51=f2num(nbi$ITEM51)/10
nbi$ITEM52=f2num(nbi$ITEM52)/10
nbi$ITEM53=as.factor(f2num(nbi$ITEM53)>3000)
nbi$ITEM54B=as.factor(f2num(nbi$ITEM54B)>0)
nbi$ITEM55B=as.factor(f2num(nbi$ITEM55B)>0)
nbi$ITEM56=as.factor(f2num(nbi$ITEM56)>0)
nbi$ITEM64=f2num(nbi$ITEM64)/10
nbi$ITEM66=f2num(nbi$ITEM66)/10
nbi$ITEM91=f2num(nbi$ITEM91)
nbi$ITEM92A[nbi$ITEM92A=="N00"]="N"
nbi$ITEM92B[nbi$ITEM92B=="N00"]="N"
nbi$ITEM92C[nbi$ITEM92C=="N00"]="N"
nbi$ITEM96=f2num(nbi$ITEM96)
nbi$ITEM103=as.factor(nbi$ITEM103)
nbi$ITEM106=f2num(nbi$ITEM106)
nbi$ITEM109=f2num(nbi$ITEM109)
nbi$ITEM111=as.factor(nbi$ITEM111)
nbi$ITEM114=f2num(nbi$ITEM114)

item.is.f=NULL
for ( j in 2:ncol(nbi) ) item.is.f=c(item.is.f,is.factor(nbi[,j]))
nbi=nbi[,c(1,1+which(item.is.f==T),1+which(item.is.f==F))]


cor.tmp=matrix(0,nrow(nbi),ncol(nbi))
for (j in 1:ncol(nbi)) cor.tmp[,j]=as.numeric(nbi[,j])
cor.mat=cor(cor.tmp)
max2=function(x) return(-sort(-x)[2])
which.max2=function(x) return(order(-x)[2])
cbind(names(nbi),names(nbi)[apply(cor.mat,2,which.max2)],apply(cor.mat,2,max2))


######################    Correlation    ########################

cor.order=order(-abs(cor.mat[1,-1]))
t.test.cor=function(r,n) {
```

```
  t.test=r*sqrt((n-2)/(1-r^2))
  p.val=2*(1-pt(abs(t.test),n-2))
  return(list(t.test=t.test,p.val=p.val))
}
cbind(names(nbi)[-1],cor.mat[1,-1],t.test.cor(cor.mat[1,-1],nrow(nbi))$p.val)[cor.order,]

##################### Multiple regression #######################

Full=lm(ITEM58~.,data=nbi)
cbind(names(nbi)[-c(1,21)],anova(Full)$"Pr(>F)"[-79])[order(anova(Full)$"Pr(>F)"[-79]),]

### AIC ###
nbi.aic=step(lm(ITEM58~1, data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=2)
m.aic=lm(ITEM58 ~ ITEM27 + ITEM43A + ITEM41 + ITEM43B + ITEM3 + ITEM91 +
ITEM22 + ITEM31 + ITEM2 + ITEM36A + ITEM30 + ITEM44B + ITEM64 + ITEM65 +
ITEM107 + ITEM36C + ITEM42B + ITEM103 + ITEM45 + ITEM106 + ITEM37 + ITEM44A +
ITEM92C + ITEM113 + ITEM20 + ITEM105 + FEDERAL + ITEM21 + ITEM109 + ITEM36D
+ ITEM26 + ITEM92A + ITEM50B + ITEM52 + ITEM108B + ITEM36B + ITEM34 + ITEM51
+ ITEM12 + ITEM92B + ITEM110 + ITEM108C + ITEM53 + ITEM111 + ITEM32 + ITEM33 +
ITEM47, data=nbi)
anova(m.aic,Full)


### BIC ###
nbi.bic=step(lm(ITEM58~1,
data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=log(nrow(nbi)))
m.bic=lm(ITEM58 ~ ITEM27 + ITEM43A + ITEM41 + ITEM91 + ITEM22 + ITEM36A +
    ITEM43B + ITEM2 + ITEM64 + ITEM65 + ITEM31 + ITEM46 + ITEM55B +
    ITEM36C + ITEM103 + ITEM45 + ITEM106 + ITEM105 + FEDERAL +
    ITEM50B + ITEM20 + ITEM52, data=nbi)
anova(m.bic,Full)

data.clean=function(x) {
  xcol=ncol(x)
  xind=NULL
  for ( j in 1:xcol ) {
    xind=c(xind,which(x[,j]=="" | x[,j]=="*"))
  }
  xind=unique(xind)
  return(x[-xind,])
}

f2num=function(x) return(as.numeric(as.character(x)))



########################### ITEM60 #######################

nbi=nbi4_60

### nbi.na=NULL
```

```
### for ( j in 1:ncol(nbi) ) nbi.na=c(nbi.na,sum(nbi[,j]=="" | nbi[,j]=="N"))

tmp.ind=which(nbi[,92]!="")
nbi.tmp=nbi[tmp.ind,92]
nbi[,92]=rep("N",nrow(nbi))
nbi[tmp.ind,92]=as.character(nbi.tmp)

tmp.ind=which(nbi[,93]!="")
nbi.tmp=nbi[tmp.ind,93]
nbi[,93]=rep("N",nrow(nbi))
nbi[tmp.ind,93]=as.character(nbi.tmp)

tmp.ind=which(nbi[,98]!="")
nbi[,98]=rep("F",nrow(nbi))
nbi[tmp.ind,98]="T"

tmp.ind=which(nbi[,108]!="")
nbi.tmp=nbi[tmp.ind,108]
nbi[,108]=rep("0",nrow(nbi))
nbi[tmp.ind,108]=as.character(nbi.tmp)

nbi=nbi[,-c(12,18:19,20:21,71:76,77:79,85:87,88:89,91,94,113,114)]

nbi=data.clean(nbi)

nbi=nbi[,-c(2:7,10:13,66,72:73,91)]

#### ITEM4??? ITEM108???

nbi$ITEM60=f2num(nbi$ITEM60)
nbi$ITEM10=as.factor(f2num(nbi$ITEM10)>3000)
nbi$ITEM11=f2num(nbi$ITEM11)/1000
nbi$ITEM19=f2num(nbi$ITEM19)
nbi$ITEM27=f2num(nbi$ITEM27)
nbi$ITEM28A=f2num(nbi$ITEM28A)
nbi$ITEM28B=f2num(nbi$ITEM28B)
nbi$ITEM29=f2num(nbi$ITEM29)
nbi$ITEM32=f2num(nbi$ITEM32)/10
nbi$ITEM34=f2num(nbi$ITEM34)
nbi$ITEM39=as.factor(f2num(nbi$ITEM39)>0)
nbi$ITEM40=as.factor(f2num(nbi$ITEM40)>0)
nbi$ITEM45=f2num(nbi$ITEM45)
nbi$ITEM46=as.factor(f2num(nbi$ITEM46)>0)
nbi$ITEM47=f2num(nbi$ITEM47)/10
nbi$ITEM48=f2num(nbi$ITEM48)/10
nbi$ITEM49=f2num(nbi$ITEM49)/10
nbi$ITEM50A=f2num(nbi$ITEM50A)/10
nbi$ITEM50B=f2num(nbi$ITEM50B)/10
nbi$ITEM51=f2num(nbi$ITEM51)/10
nbi$ITEM52=f2num(nbi$ITEM52)/10
nbi$ITEM53=as.factor(f2num(nbi$ITEM53)>3000)
```

```
nbi$ITEM54B=as.factor(f2num(nbi$ITEM54B)>0)
nbi$ITEM55B=as.factor(f2num(nbi$ITEM55B)>0)
nbi$ITEM56=as.factor(f2num(nbi$ITEM56)>0)
nbi$ITEM64=f2num(nbi$ITEM64)/10
nbi$ITEM66=f2num(nbi$ITEM66)/10
nbi$ITEM91=f2num(nbi$ITEM91)
nbi$ITEM92A[nbi$ITEM92A=="N00"]="N"
nbi$ITEM92B[nbi$ITEM92B=="N00"]="N"
nbi$ITEM92C[nbi$ITEM92C=="N00"]="N"
nbi$ITEM96=f2num(nbi$ITEM96)
nbi$ITEM103=as.factor(nbi$ITEM103)
nbi$ITEM106=f2num(nbi$ITEM106)
nbi$ITEM109=f2num(nbi$ITEM109)
nbi$ITEM111=as.factor(nbi$ITEM111)
nbi$ITEM114=f2num(nbi$ITEM114)

item.is.f=NULL
for ( j in 2:ncol(nbi) ) item.is.f=c(item.is.f,is.factor(nbi[,j]))
nbi=nbi[,c(1,1+which(item.is.f==T),1+which(item.is.f==F))]


cor.tmp=matrix(0,nrow(nbi),ncol(nbi))
for (j in 1:ncol(nbi)) cor.tmp[,j]=as.numeric(nbi[,j])
cor.mat=cor(cor.tmp)
max2=function(x) return(-sort(-x)[2])
which.max2=function(x) return(order(-x)[2])
cbind(names(nbi),names(nbi)[apply(cor.mat,2,which.max2)],apply(cor.mat,2,max2))

#####################   Correlation   ######################

cor.order=order(-abs(cor.mat[1,-1]))
t.test.cor=function(r,n) {
  t.test=r*sqrt((n-2)/(1-r^2))
  p.val=2*(1-pt(abs(t.test),n-2))
  return(list(t.test=t.test,p.val=p.val))
}
cbind(names(nbi)[-1],cor.mat[1,-1],t.test.cor(cor.mat[1,-1],nrow(nbi))$p.val)[cor.order,]


##################### Multiple regression ######################

Full=lm(ITEM60~.,data=nbi)
cbind(names(nbi)[-c(1,21)],anova(Full)$"Pr(>F)"[-79])[order(anova(Full)$"Pr(>F)"[-79]),]

### AIC ###
nbi.aic=step(lm(ITEM60~1, data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=2)
m.aic=lm(ITEM60 ~ ITEM27 + ITEM43A + ITEM41 + ITEM43B + ITEM3 + ITEM91 +
ITEM22 + ITEM31 + ITEM2 + ITEM36A + ITEM30 + ITEM44B + ITEM64 + ITEM65 +
ITEM107 + ITEM36C + ITEM42B + ITEM103 + ITEM45 + ITEM106 + ITEM37 + ITEM44A +
ITEM92C + ITEM113 + ITEM20 + ITEM105 + FEDERAL + ITEM21 + ITEM109 + ITEM36D
+ ITEM26 + ITEM92A + ITEM50B + ITEM52 + ITEM108B + ITEM36B + ITEM34 + ITEM51
```

```
+ ITEM12 + ITEM92B + ITEM110 + ITEM108C + ITEM53 + ITEM111 + ITEM32 + ITEM33 +
ITEM47, data=nbi)
anova(m.aic,Full)


### BIC ###
nbi.bic=step(lm(ITEM60~1,
data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=log(nrow(nbi)))
m.bic=lm(ITEM60 ~ ITEM27 + ITEM43A + ITEM41 + ITEM91 + ITEM22 + ITEM36A +
    ITEM43B + ITEM2 + ITEM64 + ITEM65 + ITEM31 + ITEM46 + ITEM55B +
    ITEM36C + ITEM103 + ITEM45 + ITEM106 + ITEM105 + FEDERAL +
    ITEM50B + ITEM20 + ITEM52, data=nbi)
anova(m.bic,Full)

data.clean=function(x) {
  xcol=ncol(x)
  xind=NULL
  for ( j in 1:xcol ) {
    xind=c(xind,which(x[,j]=="" | x[,j]=="*"))
  }
  xind=unique(xind)
  return(x[-xind,])
}

f2num=function(x) return(as.numeric(as.character(x)))



########################### ITEM61 ########################

nbi=nbi4_61

### nbi.na=NULL
### for ( j in 1:ncol(nbi) ) nbi.na=c(nbi.na,sum(nbi[,j]=="" | nbi[,j]=="N"))

tmp.ind=which(nbi[,92]!="")
nbi.tmp=nbi[tmp.ind,92]
nbi[,92]=rep("N",nrow(nbi))
nbi[tmp.ind,92]=as.character(nbi.tmp)

tmp.ind=which(nbi[,93]!="")
nbi.tmp=nbi[tmp.ind,93]
nbi[,93]=rep("N",nrow(nbi))
nbi[tmp.ind,93]=as.character(nbi.tmp)

tmp.ind=which(nbi[,98]!="")
nbi[,98]=rep("F",nrow(nbi))
nbi[tmp.ind,98]="T"

tmp.ind=which(nbi[,108]!="")
```

```
nbi.tmp=nbi[tmp.ind,108]
nbi[,108]=rep("0",nrow(nbi))
nbi[tmp.ind,108]=as.character(nbi.tmp)

nbi=nbi[,-c(12,18:19,20:21,71:76,77:79,85:87,88:89,91,94,113,114)]

nbi=data.clean(nbi)

nbi=nbi[,-c(2:7,10:13,66,72:73,91)]

#### ITEM4??? ITEM108???

nbi$ITEM61=f2num(nbi$ITEM61)
nbi$ITEM10=as.factor(f2num(nbi$ITEM10)>3000)
nbi$ITEM11=f2num(nbi$ITEM11)/1000
nbi$ITEM19=f2num(nbi$ITEM19)
nbi$ITEM27=f2num(nbi$ITEM27)
nbi$ITEM28A=f2num(nbi$ITEM28A)
nbi$ITEM28B=f2num(nbi$ITEM28B)
nbi$ITEM29=f2num(nbi$ITEM29)
nbi$ITEM32=f2num(nbi$ITEM32)/10
nbi$ITEM34=f2num(nbi$ITEM34)
nbi$ITEM39=as.factor(f2num(nbi$ITEM39)>0)
nbi$ITEM40=as.factor(f2num(nbi$ITEM40)>0)
nbi$ITEM45=f2num(nbi$ITEM45)
nbi$ITEM46=as.factor(f2num(nbi$ITEM46)>0)
nbi$ITEM47=f2num(nbi$ITEM47)/10
nbi$ITEM48=f2num(nbi$ITEM48)/10
nbi$ITEM49=f2num(nbi$ITEM49)/10
nbi$ITEM50A=f2num(nbi$ITEM50A)/10
nbi$ITEM50B=f2num(nbi$ITEM50B)/10
nbi$ITEM51=f2num(nbi$ITEM51)/10
nbi$ITEM52=f2num(nbi$ITEM52)/10
nbi$ITEM53=as.factor(f2num(nbi$ITEM53)>3000)
nbi$ITEM54B=as.factor(f2num(nbi$ITEM54B)>0)
nbi$ITEM55B=as.factor(f2num(nbi$ITEM55B)>0)
nbi$ITEM56=as.factor(f2num(nbi$ITEM56)>0)
nbi$ITEM64=f2num(nbi$ITEM64)/10
nbi$ITEM66=f2num(nbi$ITEM66)/10
nbi$ITEM91=f2num(nbi$ITEM91)
nbi$ITEM92A[nbi$ITEM92A=="N00"]="N"
nbi$ITEM92B[nbi$ITEM92B=="N00"]="N"
nbi$ITEM92C[nbi$ITEM92C=="N00"]="N"
nbi$ITEM96=f2num(nbi$ITEM96)
nbi$ITEM103=as.factor(nbi$ITEM103)
nbi$ITEM106=f2num(nbi$ITEM106)
nbi$ITEM109=f2num(nbi$ITEM109)
nbi$ITEM111=as.factor(nbi$ITEM111)
nbi$ITEM114=f2num(nbi$ITEM114)

item.is.f=NULL
```

```
for ( j in 2:ncol(nbi) ) item.is.f=c(item.is.f,is.factor(nbi[,j]))
nbi=nbi[,c(1,1+which(item.is.f==T),1+which(item.is.f==F))]


cor.tmp=matrix(0,nrow(nbi),ncol(nbi))
for (j in 1:ncol(nbi)) cor.tmp[,j]=as.numeric(nbi[,j])
cor.mat=cor(cor.tmp)
max2=function(x) return(-sort(-x)[2])
which.max2=function(x) return(order(-x)[2])
cbind(names(nbi),names(nbi)[apply(cor.mat,2,which.max2)],apply(cor.mat,2,max2))

######################    Correlation    #######################

cor.order=order(-abs(cor.mat[1,-1]))
t.test.cor=function(r,n) {
 t.test=r*sqrt((n-2)/(1-r^2))
 p.val=2*(1-pt(abs(t.test),n-2))
 return(list(t.test=t.test,p.val=p.val))
}
cbind(names(nbi)[-1],cor.mat[1,-1],t.test.cor(cor.mat[1,-1],nrow(nbi))$p.val)[cor.order,]


###################### Multiple regression #######################

Full=lm(ITEM61~.,data=nbi)
cbind(names(nbi)[-c(1,21)],anova(Full)$"Pr(>F)"[-79])[order(anova(Full)$"Pr(>F)"[-79]),]

### AIC ###
nbi.aic=step(lm(ITEM61~1, data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=2)
m.aic=lm(ITEM61 ~ ITEM27 + ITEM43A + ITEM41 + ITEM43B + ITEM3 + ITEM91 +
ITEM22 + ITEM31 + ITEM2 + ITEM36A + ITEM30 + ITEM44B + ITEM64 + ITEM65 +
ITEM107 + ITEM36C + ITEM42B + ITEM103 + ITEM45 + ITEM106 + ITEM37 + ITEM44A +
ITEM92C + ITEM113 + ITEM20 + ITEM105 + FEDERAL + ITEM21 + ITEM109 + ITEM36D
+ ITEM26 + ITEM92A + ITEM50B + ITEM52 + ITEM108B + ITEM36B + ITEM34 + ITEM51
+ ITEM12 + ITEM92B + ITEM110 + ITEM108C + ITEM53 + ITEM111 + ITEM32 + ITEM33 +
ITEM47, data=nbi)
anova(m.aic,Full)


### BIC ###
nbi.bic=step(lm(ITEM61~1,
data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=log(nrow(nbi)))
m.bic=lm(ITEM61 ~ ITEM27 + ITEM43A + ITEM41 + ITEM91 + ITEM22 + ITEM36A +
    ITEM43B + ITEM2 + ITEM64 + ITEM65 + ITEM31 + ITEM46 + ITEM55B +
    ITEM36C + ITEM103 + ITEM45 + ITEM106 + ITEM105 + FEDERAL +
    ITEM50B + ITEM20 + ITEM52, data=nbi)
anova(m.bic,Full)

data.clean=function(x) {
  xcol=ncol(x)
  xind=NULL
```

```
  for ( j in 1:xcol ) {
    xind=c(xind,which(x[,j]=="" | x[,j]=="*"))
  }
  xind=unique(xind)
  return(x[-xind,])
}

f2num=function(x) return(as.numeric(as.character(x)))



########################## ITEM62 ########################

nbi=nbi4_62

### nbi.na=NULL
### for ( j in 1:ncol(nbi) ) nbi.na=c(nbi.na,sum(nbi[,j]=="" | nbi[,j]=="N"))

tmp.ind=which(nbi[,92]!="")
nbi.tmp=nbi[tmp.ind,92]
nbi[,92]=rep("N",nrow(nbi))
nbi[tmp.ind,92]=as.character(nbi.tmp)

tmp.ind=which(nbi[,93]!="")
nbi.tmp=nbi[tmp.ind,93]
nbi[,93]=rep("N",nrow(nbi))
nbi[tmp.ind,93]=as.character(nbi.tmp)

tmp.ind=which(nbi[,98]!="")
nbi[,98]=rep("F",nrow(nbi))
nbi[tmp.ind,98]="T"

tmp.ind=which(nbi[,108]!="")
nbi.tmp=nbi[tmp.ind,108]
nbi[,108]=rep("0",nrow(nbi))
nbi[tmp.ind,108]=as.character(nbi.tmp)

nbi=nbi[,-c(12,18:19,20:21,71:76,77:79,85:87,88:89,91,94,113,114)]

nbi=data.clean(nbi)

nbi=nbi[,-c(2:7,10:13,66,72:73,91)]

#### ITEM4??? ITEM108???

nbi$ITEM62=f2num(nbi$ITEM62)
nbi$ITEM10=as.factor(f2num(nbi$ITEM10)>3000)
nbi$ITEM11=f2num(nbi$ITEM11)/1000
nbi$ITEM19=f2num(nbi$ITEM19)
nbi$ITEM27=f2num(nbi$ITEM27)
nbi$ITEM28A=f2num(nbi$ITEM28A)
```

```
nbi$ITEM28B=f2num(nbi$ITEM28B)
nbi$ITEM29=f2num(nbi$ITEM29)
nbi$ITEM32=f2num(nbi$ITEM32)/10
nbi$ITEM34=f2num(nbi$ITEM34)
nbi$ITEM39=as.factor(f2num(nbi$ITEM39)>0)
nbi$ITEM40=as.factor(f2num(nbi$ITEM40)>0)
nbi$ITEM45=f2num(nbi$ITEM45)
nbi$ITEM46=as.factor(f2num(nbi$ITEM46)>0)
nbi$ITEM47=f2num(nbi$ITEM47)/10
nbi$ITEM48=f2num(nbi$ITEM48)/10
nbi$ITEM49=f2num(nbi$ITEM49)/10
nbi$ITEM50A=f2num(nbi$ITEM50A)/10
nbi$ITEM50B=f2num(nbi$ITEM50B)/10
nbi$ITEM51=f2num(nbi$ITEM51)/10
nbi$ITEM52=f2num(nbi$ITEM52)/10
nbi$ITEM53=as.factor(f2num(nbi$ITEM53)>3000)
nbi$ITEM54B=as.factor(f2num(nbi$ITEM54B)>0)
nbi$ITEM55B=as.factor(f2num(nbi$ITEM55B)>0)
nbi$ITEM56=as.factor(f2num(nbi$ITEM56)>0)
nbi$ITEM64=f2num(nbi$ITEM64)/10
nbi$ITEM66=f2num(nbi$ITEM66)/10
nbi$ITEM91=f2num(nbi$ITEM91)
nbi$ITEM92A[nbi$ITEM92A=="N00"]="N"
nbi$ITEM92B[nbi$ITEM92B=="N00"]="N"
nbi$ITEM92C[nbi$ITEM92C=="N00"]="N"
nbi$ITEM96=f2num(nbi$ITEM96)
nbi$ITEM103=as.factor(nbi$ITEM103)
nbi$ITEM106=f2num(nbi$ITEM106)
nbi$ITEM109=f2num(nbi$ITEM109)
nbi$ITEM111=as.factor(nbi$ITEM111)
nbi$ITEM114=f2num(nbi$ITEM114)

item.is.f=NULL
for ( j in 2:ncol(nbi) ) item.is.f=c(item.is.f,is.factor(nbi[,j]))
nbi=nbi[,c(1,1+which(item.is.f==T),1+which(item.is.f==F))]


cor.tmp=matrix(0,nrow(nbi),ncol(nbi))
for (j in 1:ncol(nbi)) cor.tmp[,j]=as.numeric(nbi[,j])
cor.mat=cor(cor.tmp)
max2=function(x) return(-sort(-x)[2])
which.max2=function(x) return(order(-x)[2])
cbind(names(nbi),names(nbi)[apply(cor.mat,2,which.max2)],apply(cor.mat,2,max2))


######################     Correlation     #######################

cor.order=order(-abs(cor.mat[1,-1]))
t.test.cor=function(r,n) {
  t.test=r*sqrt((n-2)/(1-r^2))
  p.val=2*(1-pt(abs(t.test),n-2))
```

```
 return(list(t.test=t.test,p.val=p.val))
}
cbind(names(nbi)[-1],cor.mat[1,-1],t.test.cor(cor.mat[1,-1],nrow(nbi))$p.val)[cor.order,]


##################### Multiple regression ######################

Full=lm(ITEM62~.,data=nbi)
cbind(names(nbi)[-c(1,21)],anova(Full)$"Pr(>F)"[-79])[order(anova(Full)$"Pr(>F)"[-79]),]

### AIC ###
nbi.aic=step(lm(ITEM62~1, data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=2)
m.aic=lm(ITEM62 ~ ITEM27 + ITEM43A + ITEM41 + ITEM43B + ITEM3 + ITEM91 +
ITEM22 + ITEM31 + ITEM2 + ITEM36A + ITEM30 + ITEM44B + ITEM64 + ITEM65 +
ITEM107 + ITEM36C + ITEM42B + ITEM103 + ITEM45 + ITEM106 + ITEM37 + ITEM44A +
ITEM92C + ITEM113 + ITEM20 + ITEM105 + FEDERAL + ITEM21 + ITEM109 + ITEM36D
+ ITEM26 + ITEM92A + ITEM50B + ITEM52 + ITEM108B + ITEM36B + ITEM34 + ITEM51
+ ITEM12 + ITEM92B + ITEM110 + ITEM108C + ITEM53 + ITEM111 + ITEM32 + ITEM33 +
ITEM47, data=nbi)
anova(m.aic,Full)


### BIC ###
nbi.bic=step(lm(ITEM62~1,
data=nbi),scope=list(lower=~1,upper=Full),direction="forward",k=log(nrow(nbi)))
m.bic=lm(ITEM62 ~ ITEM27 + ITEM43A + ITEM41 + ITEM91 + ITEM22 + ITEM36A +
    ITEM43B + ITEM2 + ITEM64 + ITEM65 + ITEM31 + ITEM46 + ITEM55B +
    ITEM36C + ITEM103 + ITEM45 + ITEM106 + ITEM105 + FEDERAL +
    ITEM50B + ITEM20 + ITEM52, data=nbi)
anova(m.bic,Full)
```

Appendix C

P-Values for Explanatory Items by Region and Rating Item

Region 1

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM91	0	ITEM31	0	ITEM27	0	ITEM31	0
ITEM66	0	ITEM27	0	ITEM31	0	ITEM27	0
ITEM64	0	ITEM91	0	ITEM91	0	ITEM113	0
ITEM31	0	ITEM41	0	ITEM64	0	ITEM48	0
ITEM27	0	ITEM64	0	ITEM66	0	ITEM91	0
ITEM36A	0	ITEM66	0	ITEM36A	0	ITEM64	0
ITEM41	0	ITEM36A	0	ITEM41	0	ITEM55B	0
ITEM36B	0	ITEM36B	0	ITEM92C	0	ITEM43A	0
ITEM92C	0	ITEM92C	0	ITEM36B	0	ITEM36C	0
ITEM103	0	ITEM36C	0	ITEM36C	0	ITEM66	0
ITEM36C	0	ITEM103	0	ITEM103	0	ITEM36A	0
ITEM36D	0	ITEM92A	0	ITEM36D	0	ITEM41	0
ITEM43A	0	ITEM36D	0	ITEM92B	0	ITEM92C	0
ITEM108B	0	ITEM43B	0	ITEM45	0	ITEM56	0
ITEM106	0	ITEM37	0	ITEM46	0	ITEM63	0
ITEM46	0	ITEM46	0	ITEM92A	0	ITEM65	0
ITEM45	0	ITEM108A	0	ITEM113	0	ITEM2	0
ITEM96	0	ITEM43A	0	ITEM96	0	ITEM51	0
ITEM56	0	ITEM53	0	ITEM37	0	ITEM22	0
ITEM92B	0	ITEM44A	0	ITEM47	0	ITEM36B	0
ITEM92A	0	ITEM47	0	ITEM43A	0	ITEM21	0
ITEM44B	0	ITEM55B	0	ITEM108B	0	ITEM47	0
ITEM49	0	ITEM102	0	ITEM44A	0	ITEM92B	0
ITEM44A	2.04E-14	ITEM44B	0	ITEM102	0	ITEM42B	0
ITEM12	2.71E-14	ITEM92B	0	ITEM55B	0	ITEM52	0
ITEM37	8.88E-14	ITEM107	0	ITEM63	0	ITEM36D	0
ITEM47	3.04E-13	ITEM108B	0	ITEM65	0	ITEM102	0
ITEM42A	1.42E-12	ITEM26	0	ITEM44B	4.44E-16	ITEM114	0
ITEM114	2.19E-12	ITEM10	0	ITEM48	8.88E-16	ITEM108A	0
ITEM50A	2.12E-11	ITEM113	0	ITEM114	3.04E-13	ITEM42A	0
ITEM107	6.89E-11	ITEM96	0	ITEM26	4.67E-13	ITEM107	0
ITEM50B	1.91E-10	ITEM63	0	ITEM49	1.18E-11	ITEM104	0
ITEM113	7.00E-10	ITEM65	0	ITEM108A	1.82E-11	ITEM103	0
ITEM11	6.39E-09	ITEM45	0	ITEM42B	5.41E-11	ITEM10	0
ITEM100	9.68E-09	ITEM39	1.78E-15	ITEM107	9.22E-11	ITEM43B	0
ITEM105	3.92E-07	ITEM40	2.44E-15	ITEM53	2.25E-10	FEDERAL	0

Region 1- Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM39	6.15E-07	ITEM108C	3.77E-15	ITEM3	1.15E-09	ITEM32	0
ITEM40	6.15E-07	ITEM49	5.33E-15	ITEM39	2.94E-09	ITEM53	0
ITEM29	2.68E-06	ITEM42B	8.24E-14	ITEM40	4.07E-09	ITEM12	0
ITEM63	6.59E-06	ITEM51	3.82E-13	ITEM51	5.27E-09	ITEM50A	0
ITEM26	7.74E-06	ITEM105	1.41E-12	ITEM105	1.10E-08	ITEM3	0
ITEM109	1.02E-05	ITEM100	1.05E-10	ITEM108C	2.46E-08	ITEM30	8.15E-04
ITEM65	1.05E-05	ITEM50A	1.94E-10	ITEM56	3.00E-08	ITEM111	9.04E-04
ITEM55B	4.22E-05	ITEM42A	3.20E-10	ITEM29	4.87E-08	ITEM50B	1.00E-03
ITEM28B	5.98E-05	ITEM32	7.72E-09	ITEM43B	2.60E-07	ITEM110	1.75E-03
ITEM3	5.99E-05	ITEM54B	1.03E-08	ITEM100	5.88E-07	ITEM49	2.26E-03
ITEM53	6.75E-05	ITEM38	1.26E-08	ITEM38	9.75E-07	ITEM105	2.75E-03
ITEM43B	0.000224	ITEM50B	5.05E-08	ITEM106	9.92E-07	ITEM112	5.46E-03
ITEM102	0.000251	LONGDD	9.90E-08	ITEM32	5.04E-06	ITEM39	6.80E-03
ITEM51	0.000519	ITEM52	1.39E-05	ITEM10	1.79E-05	ITEM40	1.12E-02
ITEM30	0.000889	ITEM2	4.20E-05	ITEM12	2.12E-05	ITEM29	1.29E-02
ITEM10	0.001174	ITEM48	4.25E-05	ITEM52	2.41E-05	LATDD	1.59E-02
ITEM42B	0.002275	ITEM110	1.70E-04	ITEM54B	4.48E-05	ITEM109	1.64E-02
ITEM22	0.004141	ITEM104	6.80E-04	ITEM2	6.84E-04	ITEM45	1.65E-02
ITEM48	0.009978	ITEM114	1.54E-03	ITEM35	9.41E-04	ITEM19	3.91E-02
ITEM108C	0.010086	ITEM20	2.39E-03	ITEM50A	3.29E-03	ITEM34	3.99E-02
ITEM19	0.01086	ITEM29	4.13E-03	ITEM30	4.18E-03	ITEM92A	4.37E-02
LATDD	0.012077	ITEM34	1.09E-02	ITEM42A	0.004926	ITEM108C	4.71E-02
FEDERAL	0.02893	ITEM30	2.02E-02	LONGDD	0.008207	ITEM38	6.15E-02
ITEM112	0.030408	FEDERAL	5.26E-02	ITEM33	0.010302	ITEM46	6.62E-02
ITEM110	0.057331	ITEM22	0.066947	ITEM50B	0.012028	ITEM106	1.12E-01
ITEM32	0.087925	ITEM55A	0.085211	ITEM28B	0.033501	ITEM96	0.132279
ITEM52	0.092254	ITEM106	0.094276	ITEM112	0.045008	ITEM37	0.134088
ITEM21	0.103512	ITEM109	0.098647	ITEM110	0.063912	ITEM28A	0.161869
ITEM28A	0.115205	ITEM28A	0.110123	LATDD	0.073262	ITEM101	0.177314
ITEM35	0.133951	LATDD	0.128481	ITEM104	0.087483	ITEM35	0.218492
ITEM33	0.17113	ITEM35	0.138925	ITEM11	0.116311	ITEM11	0.302601
ITEM55A	0.361142	ITEM54A	0.156149	ITEM111	0.141584	ITEM54B	0.311664
ITEM38	0.379389	ITEM19	0.164727	FEDERAL	0.16653	ITEM54A	0.323949
ITEM108A	0.426758	ITEM111	0.203769	ITEM20	0.210698	ITEM44B	0.398666
ITEM54A	0.427963	ITEM12	0.232217	ITEM55A	0.417108	ITEM55A	0.407781
ITEM34	0.444714	ITEM112	0.256187	ITEM22	0.460686	ITEM44A	0.483524

Region 1- Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM111	0.488292	ITEM33	0.258385	ITEM21	0.489709	ITEM20	0.512718
ITEM2	0.516792	ITEM3	0.282911	ITEM34	0.553116	LONGDD	0.520563
ITEM54B	0.588268	ITEM11	0.368517	ITEM54A	0.596863	ITEM100	0.645752
ITEM101	0.707542	ITEM101	0.524147	ITEM28A	0.655166	ITEM108B	0.662902
LONGDD	0.809738	ITEM21	0.578045	ITEM101	0.781613	ITEM26	0.88589
ITEM104	0.851128	ITEM28B	0.921645	ITEM109	0.836741	ITEM33	0.92693
ITEM20	0.965738	ITEM56	0.984898	ITEM19	0.981432	ITEM28B	0.980165

Region 4

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM27	0	ITEM27	0	ITEM27	0	ITEM27	0
ITEM36A	0	ITEM36A	0	ITEM31	0	ITEM31	0
ITEM31	0	ITEM31	0	ITEM66	0	ITEM36C	0
ITEM66	0	ITEM66	0	ITEM64	0	ITEM36B	0
ITEM64	0	ITEM64	0	ITEM36A	0	ITEM66	0
ITEM36B	0	ITEM41	0	ITEM41	0	ITEM36D	0
ITEM36C	0	ITEM36C	0	ITEM36C	0	ITEM41	0
ITEM36D	0	ITEM36D	0	ITEM36B	0	ITEM36A	0
ITEM41	0	ITEM36B	0	ITEM36D	0	ITEM64	0
ITEM91	0	ITEM108A	0	ITEM108A	0	ITEM91	0
ITEM108A	0	ITEM91	0	ITEM91	0	ITEM113	0
ITEM43A	0	ITEM51	0	ITEM103	0	ITEM63	0
ITEM47	0	ITEM47	0	ITEM51	0	ITEM65	0
ITEM51	0	ITEM52	0	ITEM48	0	ITEM51	0
ITEM52	0	ITEM43A	0	ITEM52	0	ITEM56	0
LATDD	0	ITEM102	0	ITEM47	0	ITEM108A	0
ITEM37	0	ITEM48	0	ITEM107	0	ITEM102	0
ITEM102	0	ITEM107	0	ITEM102	0	ITEM55B	0
ITEM48	0	ITEM37	0	ITEM43A	0	ITEM32	0
ITEM107	0	ITEM32	0	ITEM32	0	ITEM52	0
ITEM32	0	ITEM103	0	ITEM54B	0	ITEM43A	0
ITEM103	0	ITEM54B	0	ITEM38	0	ITEM48	0
ITEM92A	0	ITEM38	0	ITEM104	0	ITEM28A	0
ITEM63	0	ITEM104	0	ITEM92C	0	ITEM12	0
ITEM65	0	ITEM28B	0	ITEM42B	0	ITEM37	0
ITEM92C	0	ITEM28A	0	ITEM28B	0	ITEM45	0
ITEM106	0	ITEM92A	0	ITEM56	0	ITEM109	0
ITEM28A	0	ITEM42B	0	ITEM55A	0	ITEM49	0
ITEM113	0	LATDD	0	ITEM54A	0	ITEM11	0
ITEM2	0	ITEM55A	0	ITEM114	0	ITEM104	0
ITEM104	0	ITEM92C	0	ITEM110	0	ITEM29	0
ITEM44B	0	ITEM54A	0	ITEM28A	0	ITEM106	0
ITEM28B	0	ITEM56	0	ITEM29	0	ITEM92A	0
ITEM30	0	ITEM106	0	ITEM100	0	ITEM114	0

Region 4 - Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM54A	0	ITEM63	0	ITEM113	0	ITEM107	0
ITEM111	0	ITEM65	0	ITEM33	0	ITEM43B	0
ITEM55A	0	ITEM29	0	ITEM12	0	ITEM110	0
ITEM108C	0	ITEM114	0	ITEM43B	0	ITEM50A	0
ITEM92B	0	ITEM33	0	ITEM34	0	ITEM50B	0
ITEM38	0	ITEM34	0	LATDD	0	ITEM42B	0
ITEM46	2.25E-12	ITEM100	0	ITEM37	0	ITEM92C	0
ITEM50B	3.25E-11	ITEM12	0	ITEM109	0	LONGDD	2.66E-15
ITEM108B	6.54E-11	ITEM110	0	ITEM42A	0	ITEM92B	2.66E-15
ITEM96	1.93E-10	ITEM113	0	ITEM63	0	ITEM54B	2.89E-15
ITEM50A	4.10E-10	ITEM19	0	ITEM65	0	ITEM40	1.91E-14
ITEM54B	1.11E-09	ITEM2	0	ITEM112	0	ITEM39	3.15E-14
ITEM43B	2.20E-09	ITEM44B	0	ITEM92B	0	ITEM34	1.41E-13
ITEM45	2.35E-09	ITEM55B	0	ITEM55B	0	ITEM3	7.98E-11
ITEM109	3.83E-09	ITEM46	0	ITEM106	0	ITEM100	1.29E-10
ITEM56	2.11E-08	ITEM30	0	ITEM11	0	ITEM2	1.60E-10
ITEM44A	3.95E-08	ITEM42A	0	ITEM92A	0	ITEM20	1.73E-10
ITEM105	5.47E-08	ITEM109	0	ITEM19	0	ITEM33	7.01E-10
ITEM19	6.26E-08	ITEM96	0	ITEM35	0	ITEM35	8.97E-10
ITEM55B	1.26E-06	ITEM111	0	ITEM49	0	ITEM46	8.01E-09
ITEM34	4.96E-06	ITEM35	0	ITEM20	0	ITEM53	5.87E-08
ITEM20	1.6E-05	ITEM44A	0	ITEM45	1.44E-12	ITEM10	7.42E-08
ITEM26	2.2E-05	ITEM53	6.66E-15	ITEM108C	1.76E-12	LATDD	1.67E-07
ITEM29	2.22E-05	ITEM108C	8.44E-15	LONGDD	2.56E-11	ITEM103	2.39E-06
ITEM35	3.35E-05	ITEM92B	2.35E-14	ITEM105	2.67E-11	ITEM28B	3.57E-06
ITEM114	9.26E-05	ITEM39	1.03E-12	ITEM10	2.19E-08	ITEM44A	9.78E-06
ITEM39	0.00011	ITEM20	1.38E-12	ITEM44B	2.13E-05	ITEM47	5.61E-05
ITEM11	0.000204	ITEM26	1.32E-11	ITEM26	9.57E-05	ITEM26	0.000246
ITEM53	0.000321	ITEM40	2.68E-11	FEDERAL	0.000558	ITEM108B	0.00033
ITEM42B	0.000542	ITEM45	5.6E-08	ITEM3	0.004702	ITEM19	0.000447
ITEM40	0.000819	ITEM43B	6.53E-08	ITEM96	0.006827	ITEM22	0.00112
ITEM100	0.001043	ITEM112	1.14E-07	ITEM22	0.00905	ITEM21	0.001639
ITEM22	0.004558	ITEM3	0.000287	ITEM39	0.012416	ITEM42A	0.001869
ITEM112	0.00492	ITEM49	0.003972	ITEM2	0.0314	ITEM30	0.026081
ITEM21	0.005222	LONGDD	0.007443	ITEM21	0.031831	ITEM44B	0.040649
ITEM12	0.005857	ITEM22	0.015651	ITEM40	0.046574	ITEM38	0.080392

Region 4							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM33	0.010203	ITEM21	0.03899	ITEM50A	0.098008	ITEM54A	0.1561
ITEM42A	0.038209	ITEM108B	0.040813	ITEM53	0.148259	ITEM111	0.192844
ITEM110	0.07857	ITEM50B	0.042143	ITEM108B	0.177139	ITEM55A	0.260997
ITEM10	0.108896	ITEM10	0.058839	ITEM50B	0.225579	ITEM108C	0.369675
FEDERAL	0.125965	ITEM11	0.129226	ITEM46	0.624434	ITEM96	0.596477
LONGDD	0.126787	ITEM50A	0.180376	ITEM111	0.665696	ITEM105	0.624987
ITEM49	0.174169	FEDERAL	0.199485	ITEM101	0.69521	ITEM112	0.715797
ITEM3	0.237969	ITEM105	0.717744	ITEM30	0.737924	ITEM101	0.840117
ITEM101	0.900811	ITEM101	0.99594	ITEM44A	0.796134	FEDERAL	0.911405

Region 5

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM27	0	ITEM27	0	ITEM27	0	ITEM27	0
ITEM91	0	ITEM66	0	ITEM31	0	ITEM31	0
ITEM36A	0	ITEM41	0	ITEM36A	0	ITEM36C	0
ITEM66	0	ITEM36A	0	ITEM41	0	ITEM36B	0
ITEM41	0	ITEM31	0	ITEM66	0	ITEM66	0
ITEM31	0	ITEM91	0	ITEM91	0	ITEM36D	0
ITEM36D	0	ITEM64	0	ITEM64	0	ITEM41	0
ITEM36C	0	ITEM108A	0	ITEM36C	0	ITEM36A	0
ITEM64	0	ITEM36D	0	ITEM36D	0	ITEM64	0
ITEM108A	0	ITEM36C	0	ITEM108A	0	ITEM91	0
ITEM36B	0	ITEM43A	0	ITEM36B	0	ITEM113	0
ITEM43A	0	ITEM92A	0	ITEM106	0	ITEM63	0
ITEM108C	0	ITEM37	0	ITEM37	0	ITEM65	0
ITEM37	0	ITEM36B	0	ITEM43A	0	ITEM51	0
ITEM92A	0	ITEM108C	0	ITEM92A	0	ITEM56	0
ITEM55B	0	ITEM53	0	ITEM102	0	ITEM108A	0
ITEM103	0	ITEM10	0	ITEM108C	0	ITEM102	0
ITEM2	0	ITEM102	0	ITEM63	0	ITEM55B	0
ITEM30	0	ITEM106	0	ITEM65	0	ITEM32	0
ITEM53	0	ITEM107	0	ITEM107	0	ITEM52	0
ITEM10	0	ITEM43B	0	ITEM51	0	ITEM43A	0
ITEM56	0	ITEM30	0	ITEM47	0	ITEM48	0
ITEM102	0	ITEM103	0	ITEM113	0	ITEM28A	0
ITEM42A	0	ITEM51	0	ITEM30	0	ITEM12	0
ITEM106	0	ITEM47	0	ITEM55B	0	ITEM37	0
ITEM92C	0	ITEM38	0	ITEM38	0	ITEM45	0
ITEM63	0	ITEM113	0	ITEM53	0	ITEM109	0
ITEM65	0	ITEM63	0	ITEM10	0	ITEM49	0
ITEM107	0	ITEM65	0	ITEM43B	0	ITEM11	0
ITEM110	0	ITEM55B	0	ITEM52	0	ITEM104	0
ITEM50B	0	ITEM44A	0	ITEM32	0	ITEM29	0
ITEM50A	0	ITEM44B	0	ITEM103	0	ITEM106	0
ITEM12	0	ITEM34	0	ITEM48	0	ITEM92A	0
ITEM44A	0	ITEM46	0	ITEM56	0	ITEM114	0
ITEM44B	0	ITEM92C	0	ITEM44A	0	ITEM107	0
ITEM51	0	ITEM52	0	ITEM34	0	ITEM43B	0

Region 5 - Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM113	0	ITEM2	0	ITEM44B	0	ITEM110	0
ITEM46	0	ITEM32	0	ITEM46	0	ITEM50A	0
ITEM45	0	ITEM48	0	ITEM92C	0	ITEM50B	0
ITEM47	0	ITEM42A	0	ITEM28A	0	ITEM42B	0
ITEM92B	2.22E-16	ITEM40	2.22E-16	ITEM42A	0	ITEM92C	0
ITEM38	2.22E-16	ITEM39	2.22E-16	ITEM109	0	LONGDD	2.66E-15
ITEM54B	4.44E-16	ITEM50A	9.33E-15	ITEM92B	0	ITEM92B	2.66E-15
ITEM21	3.11E-15	ITEM50B	1.47E-13	ITEM2	0	ITEM54B	2.89E-15
ITEM100	2.75E-14	ITEM92B	8.26E-13	ITEM40	1.14E-09	ITEM40	1.91E-14
ITEM22	6.57E-14	ITEM108B	1.23E-12	ITEM39	1.46E-09	ITEM39	3.15E-14
ITEM29	9.90E-14	ITEM28A	2.93E-12	ITEM20	4.75E-09	ITEM34	1.41E-13
ITEM114	2.65E-13	ITEM56	4.23E-12	ITEM26	9.95E-09	ITEM3	7.98E-11
ITEM33	4.87E-12	ITEM110	1.91E-11	ITEM50A	1.43E-08	ITEM100	1.29E-10
ITEM48	7.23E-12	ITEM26	2.23E-10	ITEM19	3.94E-08	ITEM2	1.60E-10
ITEM42B	1.07E-11	LATDD	3.01E-09	ITEM50B	4.27E-08	ITEM20	1.73E-10
ITEM43B	3.58E-10	ITEM21	2.14E-08	ITEM112	4.88E-07	ITEM33	7.01E-10
ITEM104	7.68E-09	ITEM22	1.48E-07	ITEM104	2.74E-06	ITEM35	8.97E-10
ITEM40	1.26E-08	ITEM11	4.40E-06	ITEM54A	6.03E-06	ITEM46	8.01E-09
ITEM39	1.61E-08	ITEM33	4.99E-06	ITEM55A	9.37E-06	ITEM53	5.87E-08
ITEM34	2.28E-08	ITEM111	1.64E-05	ITEM108B	2.38E-05	ITEM10	7.42E-08
ITEM28B	3.58E-08	ITEM96	5.95E-05	ITEM42B	2.59E-05	LATDD	1.67E-07
ITEM11	8.40E-08	ITEM12	1.08E-04	ITEM111	0.000132	ITEM103	2.39E-06
ITEM49	2.42E-06	ITEM112	1.09E-04	ITEM110	0.000563	ITEM28B	3.57E-06
ITEM111	2.23E-05	ITEM109	1.09E-04	ITEM3	0.000672	ITEM44A	9.78E-06
ITEM112	0.000706	ITEM45	0.000156	ITEM12	0.001592	ITEM47	5.61E-05
ITEM26	0.001566	LONGDD	0.000425	ITEM11	0.00263	ITEM26	0.000246
ITEM52	0.00204	ITEM29	0.001242	ITEM29	0.005294	ITEM108B	0.00033
ITEM3	0.005219	ITEM54A	0.002115	ITEM33	0.007135	ITEM19	0.000447
ITEM96	0.008561	ITEM114	0.002386	FEDERAL	0.008043	ITEM22	0.00112
LATDD	0.014439	ITEM55A	0.002823	ITEM114	0.026623	ITEM21	0.001639
ITEM35	0.015327	ITEM49	0.004611	LATDD	0.028819	ITEM42A	0.001869
ITEM109	0.018716	ITEM104	0.014815	ITEM54B	0.037579	ITEM30	0.026081
ITEM19	0.024443	ITEM19	0.016743	ITEM100	0.086793	ITEM44B	0.040649
ITEM28A	0.037733	ITEM100	0.018509	ITEM22	0.09515	ITEM38	0.080392
FEDERAL	0.047534	ITEM42B	0.068952	ITEM45	0.125412	ITEM54A	0.1561
ITEM55A	0.137702	ITEM3	0.214704	ITEM28B	0.131291	ITEM111	0.192844

Region 5 - Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM54A	0.15022	ITEM28B	0.247892	ITEM21	0.255447	ITEM55A	0.260997
LONGDD	0.212769	ITEM35	0.270166	ITEM96	0.255769	ITEM108C	0.369675
ITEM32	0.27493	ITEM54B	0.411827	ITEM49	0.341472	ITEM96	0.596477
ITEM20	0.37263	ITEM105	0.459961	ITEM105	0.46031	ITEM105	0.624987
ITEM101	0.55881	ITEM101	0.50718	LONGDD	0.529612	ITEM112	0.715797
ITEM108B	0.824434	ITEM20	0.531526	ITEM101	0.669168	ITEM101	0.840117
ITEM105	0.912238	FEDERAL	0.797443	ITEM35	0.912418	FEDERAL	0.911405

Region 9

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM27	0	ITEM27	0	ITEM27	0	ITEM27	0
ITEM91	0	ITEM66	0	ITEM31	0	ITEM31	0
ITEM36A	0	ITEM41	0	ITEM36A	0	ITEM36C	0
ITEM66	0	ITEM36A	0	ITEM41	0	ITEM36B	0
ITEM41	0	ITEM31	0	ITEM66	0	ITEM66	0
ITEM31	0	ITEM91	0	ITEM91	0	ITEM36D	0
ITEM36D	0	ITEM64	0	ITEM64	0	ITEM41	0
ITEM36C	0	ITEM108A	0	ITEM36C	0	ITEM36A	0
ITEM64	0	ITEM36D	0	ITEM36D	0	ITEM64	0
ITEM108A	0	ITEM36C	0	ITEM108A	0	ITEM91	0
ITEM36B	0	ITEM43A	0	ITEM36B	0	ITEM113	0
ITEM43A	0	ITEM92A	0	ITEM106	0	ITEM63	0
ITEM108C	0	ITEM37	0	ITEM37	0	ITEM65	0
ITEM37	0	ITEM36B	0	ITEM43A	0	ITEM51	0
ITEM92A	0	ITEM108C	0	ITEM92A	0	ITEM56	0
ITEM55B	0	ITEM53	0	ITEM102	0	ITEM108A	0
ITEM103	0	ITEM10	0	ITEM108C	0	ITEM102	0
ITEM2	0	ITEM102	0	ITEM63	0	ITEM55B	0
ITEM30	0	ITEM106	0	ITEM65	0	ITEM32	0
ITEM53	0	ITEM107	0	ITEM107	0	ITEM52	0
ITEM10	0	ITEM43B	0	ITEM51	0	ITEM43A	0
ITEM56	0	ITEM30	0	ITEM47	0	ITEM48	0
ITEM102	0	ITEM103	0	ITEM113	0	ITEM28A	0
ITEM42A	0	ITEM51	0	ITEM30	0	ITEM12	0
ITEM106	0	ITEM47	0	ITEM55B	0	ITEM37	0
ITEM92C	0	ITEM38	0	ITEM38	0	ITEM45	0
ITEM63	0	ITEM113	0	ITEM53	0	ITEM109	0
ITEM65	0	ITEM63	0	ITEM10	0	ITEM49	0
ITEM107	0	ITEM65	0	ITEM43B	0	ITEM11	0
ITEM110	0	ITEM55B	0	ITEM52	0	ITEM104	0
ITEM50B	0	ITEM44A	0	ITEM32	0	ITEM29	0
ITEM50A	0	ITEM44B	0	ITEM103	0	ITEM106	0
ITEM12	0	ITEM34	0	ITEM48	0	ITEM92A	0
ITEM44A	0	ITEM46	0	ITEM56	0	ITEM114	0
ITEM44B	0	ITEM92C	0	ITEM44A	0	ITEM107	0
ITEM51	0	ITEM52	0	ITEM34	0	ITEM43B	0

Region 9 - Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM113	0	ITEM2	0	ITEM44B	0	ITEM110	0
ITEM46	0	ITEM32	0	ITEM46	0	ITEM50A	0
ITEM45	0	ITEM48	0	ITEM92C	0	ITEM50B	0
ITEM47	0	ITEM42A	0	ITEM28A	0	ITEM42B	0
ITEM92B	2.22E-16	ITEM40	2.22E-16	ITEM42A	0	ITEM92C	0
ITEM38	2.22E-16	ITEM39	2.22E-16	ITEM109	0	LONGDD	2.66E-15
ITEM54B	4.44E-16	ITEM50A	9.33E-15	ITEM92B	0	ITEM92B	2.66E-15
ITEM21	3.11E-15	ITEM50B	1.47E-13	ITEM2	0	ITEM54B	2.89E-15
ITEM100	2.75E-14	ITEM92B	8.26E-13	ITEM40	1.14E-09	ITEM40	1.91E-14
ITEM22	6.57E-14	ITEM108B	1.23E-12	ITEM39	1.46E-09	ITEM39	3.15E-14
ITEM29	9.90E-14	ITEM28A	2.93E-12	ITEM20	4.75E-09	ITEM34	1.41E-13
ITEM114	2.65E-13	ITEM56	4.23E-12	ITEM26	9.95E-09	ITEM3	7.98E-11
ITEM33	4.87E-12	ITEM110	1.91E-11	ITEM50A	1.43E-08	ITEM100	1.29E-10
ITEM48	7.23E-12	ITEM26	2.23E-10	ITEM19	3.94E-08	ITEM2	1.60E-10
ITEM42B	1.07E-11	LATDD	3.01E-09	ITEM50B	4.27E-08	ITEM20	1.73E-10
ITEM43B	3.58E-10	ITEM21	2.14E-08	ITEM112	4.88E-07	ITEM33	7.01E-10
ITEM104	7.68E-09	ITEM22	1.48E-07	ITEM104	2.74E-06	ITEM35	8.97E-10
ITEM40	1.26E-08	ITEM11	4.40E-06	ITEM54A	6.03E-06	ITEM46	8.01E-09
ITEM39	1.61E-08	ITEM33	4.99E-06	ITEM55A	9.37E-06	ITEM53	5.87E-08
ITEM34	2.28E-08	ITEM111	1.64E-05	ITEM108B	2.38E-05	ITEM10	7.42E-08
ITEM28B	3.58E-08	ITEM96	5.95E-05	ITEM42B	2.59E-05	LATDD	1.67E-07
ITEM11	8.40E-08	ITEM12	1.08E-04	ITEM111	0.000132	ITEM103	2.39E-06
ITEM49	2.42E-06	ITEM112	1.09E-04	ITEM110	0.000563	ITEM28B	3.57E-06
ITEM111	2.23E-05	ITEM109	1.09E-04	ITEM3	0.000672	ITEM44A	9.78E-06
ITEM112	7.06E-04	ITEM45	0.000156	ITEM12	0.001592	ITEM47	5.61E-05
ITEM26	1.57E-03	LONGDD	0.000425	ITEM11	0.00263	ITEM26	0.000246
ITEM52	2.04E-03	ITEM29	0.001242	ITEM29	0.005294	ITEM108B	0.00033
ITEM3	5.22E-03	ITEM54A	0.002115	ITEM33	0.007135	ITEM19	0.000447
ITEM96	8.56E-03	ITEM114	0.002386	FEDERAL	0.008043	ITEM22	0.00112
LATDD	1.44E-02	ITEM55A	0.002823	ITEM114	0.026623	ITEM21	0.001639
ITEM35	1.53E-02	ITEM49	0.004611	LATDD	0.028819	ITEM42A	0.001869
ITEM109	1.87E-02	ITEM104	0.014815	ITEM54B	0.037579	ITEM30	0.026081
ITEM19	2.44E-02	ITEM19	0.016743	ITEM100	0.086793	ITEM44B	0.040649
ITEM28A	3.77E-02	ITEM100	0.018509	ITEM22	0.09515	ITEM38	0.080392
FEDERAL	4.75E-02	ITEM42B	0.068952	ITEM45	0.125412	ITEM54A	0.1561
ITEM55A	1.38E-01	ITEM3	0.214704	ITEM28B	0.131291	ITEM111	0.192844

Region 9 - Continued							
Item 58	**P-Value**	**Item 59**	**P-Value**	**Item 60**	**P-Value**	**Item 61**	**P-Value**
ITEM54A	1.50E-01	ITEM28B	0.247892	ITEM21	0.255447	ITEM55A	0.260997
LONGDD	2.13E-01	ITEM35	0.270166	ITEM96	0.255769	ITEM108C	0.369675
ITEM32	2.75E-01	ITEM54B	0.411827	ITEM49	0.341472	ITEM96	0.596477
ITEM20	3.73E-01	ITEM105	0.459961	ITEM105	0.46031	ITEM105	0.624987
ITEM101	5.59E-01	ITEM101	0.50718	LONGDD	0.529612	ITEM112	0.715797
ITEM108B	8.24E-01	ITEM20	0.531526	ITEM101	0.669168	ITEM101	0.840117
ITEM105	9.12E-01	FEDERAL	0.797443	ITEM35	0.912418	FEDERAL	0.911405

Region 10

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM27	0	ITEM27	0	ITEM27	0	ITEM36A	0
ITEM64	0	ITEM31	0	ITEM31	0	ITEM36B	0
ITEM31	0	ITEM64	0	ITEM64	0	ITEM36C	0
ITEM66	0	ITEM66	0	ITEM36A	0	ITEM36D	0
ITEM107	0	ITEM41	0	ITEM91	0	ITEM27	0
ITEM92A	0	ITEM36A	0	ITEM66	0	ITEM91	0
ITEM43A	0	ITEM91	0	ITEM41	0	ITEM30	0
ITEM91	0	ITEM43A	0	ITEM36B	0	ITEM108A	0
ITEM41	0	ITEM92A	0	ITEM36C	0	ITEM48	0
ITEM36A	0	ITEM46	0	ITEM36D	0	ITEM32	0
ITEM46	0	ITEM37	0	ITEM108A	0	ITEM31	0
ITEM12	0	ITEM43B	0	ITEM47	0	ITEM47	0
ITEM37	0	ITEM36B	0	ITEM103	0	ITEM51	0
ITEM104	0	ITEM44B	0	ITEM51	0	ITEM66	0
ITEM55B	0	ITEM53	0	ITEM32	0	ITEM64	0
ITEM11	0	ITEM36D	0	ITEM37	0	ITEM102	0
ITEM45	0	ITEM106	0	ITEM52	0	ITEM55B	0
ITEM44B	0	ITEM63	0	ITEM63	0	ITEM52	0
ITEM49	0	ITEM65	0	ITEM65	0	LATDD	0
ITEM110	0	ITEM44A	0	ITEM92B	0	ITEM111	0
ITEM54B	0	ITEM103	0	ITEM102	0	ITEM110	0
ITEM38	0	ITEM108A	0	ITEM48	0	ITEM54B	0
ITEM36B	0	ITEM10	0	ITEM106	0	ITEM49	0
ITEM100	0	ITEM36C	0	ITEM30	0	ITEM2	0
ITEM108C	0	ITEM107	0	ITEM43A	0	ITEM3	0
ITEM92B	0	ITEM108C	0	ITEM108C	0	ITEM104	0
ITEM108B	0	ITEM49	0	ITEM42B	0	ITEM29	0
ITEM44A	0	ITEM45	0	ITEM92A	0	ITEM28A	0
ITEM65	0	ITEM47	0	ITEM38	0	ITEM108C	0
ITEM63	0	ITEM92B	0	ITEM54B	0	ITEM114	0
ITEM114	0	ITEM51	0	ITEM113	0	ITEM108B	0
ITEM28B	0	ITEM39	0	ITEM45	0	ITEM12	0
ITEM53	0	ITEM40	0	ITEM108B	0	ITEM41	0
ITEM106	0	ITEM11	0	ITEM46	0	ITEM42B	0
ITEM29	0	ITEM55B	0	ITEM44B	0	ITEM43B	0
ITEM103	0	ITEM30	0	ITEM56	0	ITEM92C	0

Region 10 – Continued

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM36D	0	ITEM92C	0	ITEM28A	0	ITEM20	0
ITEM10	0	ITEM28B	0	ITEM55A	0	ITEM39	0
ITEM39	0	ITEM12	0	ITEM54A	0	ITEM28B	0
ITEM40	0	ITEM2	6.66E-16	ITEM43B	0	ITEM40	0
ITEM43B	0.00E+00	ITEM54B	8.88E-16	ITEM104	0	ITEM56	0
ITEM42B	0.00E+00	ITEM110	1.33E-15	ITEM44A	0	ITEM65	3.12E-11
ITEM56	0.00E+00	ITEM108B	1.38E-14	ITEM29	4.44E-16	ITEM19	3.30E-11
ITEM30	0.00E+00	ITEM52	3.42E-14	ITEM114	6.66E-16	ITEM63	4.57E-11
ITEM2	4.44E-15	ITEM114	3.79E-12	ITEM110	8.66E-15	ITEM33	5.94E-11
ITEM92C	2.53E-14	ITEM104	5.59E-12	ITEM28B	1.24E-14	ITEM100	7.82E-11
ITEM55A	3.55E-14	ITEM38	1.23E-11	ITEM40	1.11E-13	ITEM38	6.51E-10
ITEM35	1.91E-13	ITEM100	6.78E-10	ITEM39	1.87E-13	ITEM113	8.51E-08
LATDD	1.95E-12	ITEM29	8.69E-10	ITEM55B	6.55E-13	ITEM53	9.28E-08
ITEM54A	2.26E-12	ITEM32	2.31E-09	ITEM19	1.72E-12	ITEM103	4.04E-07
ITEM109	2.83E-11	ITEM113	6.00E-07	ITEM10	2.06E-10	ITEM109	4.94E-07
ITEM48	9.02E-11	ITEM55A	7.49E-07	ITEM33	3.02E-10	ITEM54A	4.97E-06
ITEM3	2.80E-10	ITEM54A	1.81E-06	ITEM53	3.53E-10	FEDERAL	5.28E-06
ITEM113	4.33E-10	ITEM20	1.89E-06	ITEM2	1.36E-09	ITEM55A	7.46E-06
ITEM42A	4.96E-09	ITEM3	6.89E-06	FEDERAL	3.29E-07	ITEM22	8.94E-06
ITEM33	7.67E-08	ITEM56	1.05E-05	ITEM34	4.97E-07	ITEM10	2.25E-05
ITEM36C	1.11E-07	ITEM35	1.20E-05	ITEM3	3.53E-06	ITEM42A	2.39E-05
ITEM108A	1.83E-07	ITEM50A	1.51E-05	ITEM49	9.29E-06	ITEM35	5.56E-05
ITEM102	3.24E-07	LATDD	1.87E-05	LATDD	2.54E-05	ITEM50B	9.57E-05
ITEM50A	1.74E-06	ITEM19	7.88E-05	ITEM12	6.91E-05	ITEM50A	2.48E-04
ITEM47	2.03E-06	ITEM42A	9.80E-05	ITEM92C	0.000134	ITEM45	4.28E-04
ITEM32	7.16E-05	FEDERAL	0.000695	ITEM109	0.000863	ITEM21	0.000958
ITEM19	9.95E-05	ITEM42B	0.001228	ITEM100	0.002791	ITEM44B	0.001828
ITEM26	0.000234	ITEM22	0.001374	ITEM26	0.006045	ITEM37	0.002057
ITEM21	0.000437	ITEM102	0.002665	ITEM96	0.008433	ITEM92A	0.006957
ITEM20	0.000695	ITEM21	0.003147	ITEM111	0.012518	ITEM106	0.018267
ITEM22	0.001186	ITEM96	0.003284	ITEM50B	0.015765	ITEM92B	0.036011
ITEM96	0.001902	ITEM28A	0.008262	ITEM11	0.044539	ITEM105	0.042168
ITEM50B	0.004141	ITEM50B	0.008945	ITEM105	0.056889	ITEM112	0.069881
ITEM28A	0.005043	ITEM34	0.016988	ITEM22	0.088977	ITEM34	0.07179
ITEM51	0.016438	ITEM105	0.034216	ITEM42A	0.094654	ITEM96	0.080091
ITEM105	0.033574	ITEM111	0.050981	ITEM35	0.135997	ITEM107	0.100217

Region 10 – Continued

Item 58	P-Value	Item 59	P-Value	Item 60	P-Value	Item 61	P-Value
ITEM111	0.044069	LONGDD	0.052187	LONGDD	0.162362	ITEM101	0.205797
ITEM101	0.475917	ITEM48	0.152506	ITEM20	0.220264	LONGDD	0.268823
ITEM34	0.524937	ITEM26	0.187279	ITEM112	0.233046	ITEM26	0.301606
ITEM52	0.831702	ITEM109	0.212157	ITEM50A	0.270569	ITEM11	0.339962
LONGDD	0.845566	ITEM33	0.272032	ITEM21	0.788631	ITEM44A	0.505442
FEDERAL	0.933246	ITEM112	0.383645	ITEM101	0.909305	ITEM46	0.520304
ITEM112	0.988642	ITEM101	0.783154	ITEM107	0.92644	ITEM43A	0.592956

Appendix D

Correlation Matrices for Top 20 Items by Region

Region 1

	ITEM106	ITEM107	ITEM108A	ITEM108C	ITEM113	ITEM2	ITEM27	ITEM3	ITEM31	ITEM36A	ITEM36D
ITEM106	1	0.067549	0.088885	0.012739	0.066043	-0.04375	-0.41093	-0.0306	-0.05997	0.099355	0.041721
ITEM107	0.067549	1	0.535082	0.569084	-0.53557	0.062182	-0.62135	0.083465	-0.6642	-0.39124	-0.56957
ITEM108A	0.088885	0.535082	1	0.59391	-0.10518	-0.06139	-0.31412	-0.17685	-0.2324	-0.16705	-0.18171
ITEM108C	0.012739	0.569084	0.59391	1	-0.16866	0.022476	-0.37187	0.268988	-0.26576	-0.20013	-0.17506
ITEM113	0.066043	-0.53557	-0.10518	-0.16866	1	-0.03953	0.525091	-0.12159	0.6181	0.411196	0.636399
ITEM2	-0.04375	0.062182	-0.06139	0.022476	-0.03953	1	-0.14519	0.357421	-0.178	-0.11453	-0.03229
ITEM27	-0.41093	-0.62135	-0.31412	-0.37187	0.525091	-0.14519	1	-0.2642	0.857967	0.649537	0.660729
ITEM3	-0.0306	0.083465	-0.17685	0.268988	-0.12159	0.357421	-0.2642	1	-0.28215	-0.24246	-0.0931
ITEM31	-0.05997	-0.6642	-0.2324	-0.26576	0.6181	-0.178	0.857967	-0.28215	1	0.750146	0.770434
ITEM36A	0.099355	-0.39124	-0.16705	-0.20013	0.411196	-0.11453	0.649537	-0.24246	0.750146	1	0.75397
ITEM36D	0.041721	-0.56957	-0.18171	-0.17506	0.636399	-0.03229	0.660729	-0.0931	0.770434	0.75397	1
ITEM37	-0.24051	-0.52171	-0.23799	-0.47856	0.262575	-0.07113	0.632823	-0.27041	0.446244	0.314598	0.323799
ITEM41	0.107932	0.540728	0.203618	0.212558	-0.43342	0.111301	-0.71037	0.217059	-0.75386	-0.60777	-0.59245
ITEM43A	0.069932	0.363961	0.153756	0.239705	0.097343	-0.04536	0.01873	0.003923	0.066654	0.068651	0.025443
ITEM43B	0.150474	0.67762	0.399397	0.52839	-0.36652	0.060578	-0.60704	0.1222	-0.49607	-0.3052	-0.37443
ITEM45	0.079313	-0.39137	-0.19858	-0.19761	0.345339	-0.04054	0.230383	-0.02799	0.305184	0.142376	0.329489
ITEM48	0.00908	-0.51436	-0.24291	-0.22144	0.493273	-0.13796	0.450471	-0.14109	0.558512	0.323419	0.483005
ITEM55B	-0.09607	-0.51268	-0.13659	-0.41694	0.602803	-0.19588	0.592857	-0.48244	0.610622	0.406651	0.45909
ITEM91	-0.08433	-0.25999	-0.13295	-0.11532	0.197109	-0.1469	0.502213	-0.20292	0.489231	0.47281	0.357108
ITEM92B	0.050303	0.229742	0.045508	0.179973	-0.54688	0.034961	-0.40462	0.175387	-0.31976	-0.28699	-0.34077
ITEM92C	0.048358	0.020772	0.085384	0.061302	-0.03331	-0.08149	-0.23146	-0.01448	-0.15716	-0.25456	-0.11419

Region 1 - Continued

	ITEM37	ITEM41	ITEM43A	ITEM43B	ITEM45	ITEM48	ITEM55B	ITEM91	ITEM92B	ITEM92C
ITEM106	-0.24051	0.107932	0.069932	0.150474	0.079313	0.00908	-0.09607	-0.08433	0.050303	0.048358
ITEM107	-0.52171	0.540728	0.363961	0.67762	-0.39137	-0.51436	-0.51268	-0.25999	0.229742	0.020772
ITEM108A	-0.23799	0.203618	0.153756	0.399397	-0.19858	-0.24291	-0.13659	-0.13295	0.045508	0.085384
ITEM108C	-0.47856	0.212558	0.239705	0.52839	-0.19761	-0.22144	-0.41694	-0.11532	0.179973	0.061302
ITEM113	0.262575	-0.43342	0.097343	-0.36652	0.345339	0.493273	0.602803	0.197109	-0.54688	-0.03331
ITEM2	-0.07113	0.111301	-0.04536	0.060578	-0.04054	-0.13796	-0.19588	-0.1469	0.034961	-0.08149
ITEM27	0.632823	-0.71037	0.01873	-0.60704	0.230383	0.450471	0.592857	0.502213	-0.40462	-0.23146
ITEM3	-0.27041	0.217059	0.003923	0.1222	-0.02799	-0.14109	-0.48244	-0.20292	0.175387	-0.01448
ITEM31	0.446244	-0.75386	0.066654	-0.49607	0.305184	0.558512	0.610622	0.489231	-0.31976	-0.15716
ITEM36A	0.314598	-0.60777	0.068651	-0.3052	0.142376	0.323419	0.406651	0.47281	-0.28699	-0.25456
ITEM36D	0.323799	-0.59245	0.025443	-0.37443	0.329489	0.483005	0.45909	0.357108	-0.34077	-0.11419
ITEM37	1	-0.5671	-0.27295	-0.75501	-0.0173	-0.03313	0.493636	0.422607	-0.41838	-0.27879
ITEM41	-0.5671	1	0.087653	0.486044	-0.0984	-0.26653	-0.52884	-0.70011	0.302938	0.367467
ITEM43A	-0.27295	0.087653	1	0.177377	0.10859	0.149277	-0.13754	-0.04395	0.00325	0.052106
ITEM43B	-0.75501	0.486044	0.177377	1	-0.10512	0.000258	-0.47753	-0.35161	0.426065	0.226589
ITEM45	-0.0173	-0.0984	0.10859	-0.10512	1	0.618547	0.164774	-0.1309	0.102613	0.28122
ITEM48	-0.03313	-0.26653	0.149277	0.000258	0.618547	1	0.332763	0.030508	0.017085	0.243752
ITEM55B	0.493636	-0.52884	-0.13754	-0.47753	0.164774	0.332763	1	0.330001	-0.46433	-0.13088
ITEM91	0.422607	-0.70011	-0.04395	-0.35161	-0.1309	0.030508	0.330001	1	-0.28188	-0.69382
ITEM92B	-0.41838	0.302938	0.00325	0.426065	0.102613	0.017085	-0.46433	-0.28188	1	0.205276
ITEM92C	-0.27879	0.367467	0.052106	0.226589	0.28122	0.243752	-0.13088	-0.69382	0.205276	1

Region 4

	ITEM106	ITEM107	ITEM108A	ITEM108C	ITEM2	ITEM27	ITEM3	ITEM31	ITEM36A	ITEM36D
ITEM106	1	-0.1254326	-0.0110048	0.052176746	-0.0292016	-0.2861342	-0.076865	0.0872101	0.0773128	0.0840798
ITEM107	-0.1254325	1	0.8254644	-0.200875892	0.2812171	-0.5202654	0.0536346	-0.8528474	-0.7083057	-0.6913114
ITEM108A	-0.0110048	0.8254644	1	-0.040565894	0.1478557	-0.6888503	0.0132358	-0.8593133	-0.7653359	-0.7013569
ITEM108C	0.0521767	-0.2008759	-0.0405659	1	-0.2510153	0.0575749	-0.011526	0.0301345	0.0262631	-0.0102984
ITEM2	-0.0292016	0.2812172	0.1478557	-0.251015312	1	-0.0235151	0.0350799	-0.1218568	-0.160194	-0.1418902
ITEM27	-0.2861342	-0.5202654	-0.6888503	0.057574864	-0.0235151	1	-0.0313649	0.7723342	0.831896	0.6917706
ITEM3	-0.076865	0.0536346	0.0132358	-0.011525951	0.0350799	-0.0313649	1	-0.0543212	-0.0732069	-0.1083574
ITEM31	0.0872101	-0.8528474	-0.8593133	0.030134465	-0.1218568	0.7723342	-0.0543212	1	0.8691154	0.8190246
ITEM36A	0.0773128	-0.7083057	-0.7653359	0.026263104	-0.160194	0.831896	-0.0732069	0.8691154	1	0.8877304
ITEM36D	0.0840798	-0.6913114	-0.7013569	-0.010298358	-0.1418902	0.6917706	-0.1083574	0.8190246	0.8877304	1
ITEM37	-0.1301175	-0.2107793	-0.3662618	-0.067079632	0.4042509	0.3418298	0.0589588	0.2284335	0.2278444	0.2284383
ITEM41	-0.1221676	0.8060505	0.736564	-0.115978601	0.1753217	-0.6614568	0.0728568	-0.903605	-0.7724653	-0.709998
ITEM43A	-0.1169348	-0.2315252	-0.4426148	-0.235926532	0.1270368	0.6504821	0.0209989	0.5391906	0.4944016	0.4291178
ITEM43B	-0.1835067	0.3934533	0.323131	-0.272563725	-0.1271776	-0.2772775	-0.0309104	-0.4323478	-0.26389	-0.3231784
ITEM45	0.1023667	-0.3212568	-0.3053408	0.122655113	-0.1624466	0.0680847	-0.0868817	0.2203958	0.174518	0.161539
ITEM48	0.0520891	-0.6671827	-0.7543519	-0.088176683	-0.0954278	0.5170462	-0.077589	0.7698994	0.622784	0.6296259
ITEM55B	0.1566138	-0.318152	-0.3297359	0.157508359	0.3237726	0.3035638	-0.1386794	0.4315045	0.2820259	0.3626731
ITEM91	0.0131603	-0.475265	-0.5188051	-0.243149416	-0.0117736	0.5205166	0.1293244	0.6504145	0.5788049	0.5516007
ITEM92B	0.0280779	-0.0544566	-0.0261177	0.014431769	-0.1145417	-0.1145838	0.0226631	-0.0677645	-0.0847161	-0.1286574
ITEM92C	-0.0599854	0.1833165	0.1787795	-0.008896844	-0.1420204	-0.2801802	-0.134611	-0.304926	-0.2404074	-0.2464428

Region 4 - Continued

	ITEM37	ITEM41	ITEM43A	ITEM43B	ITEM45	ITEM48	ITEM55B	ITEM91	ITEM92B	ITEM92C
ITEM106	-0.1301175	-0.1221676	-0.1169348	-0.1835067	0.1023667	0.0520891	0.1566138	0.0131603	0.0280779	-0.0599854
ITEM107	-0.2107793	0.8060505	-0.2315252	0.3934533	-0.3212568	-0.6671827	-0.318152	-0.475265	-0.0544566	0.1833165
ITEM108A	-0.3662618	0.736564	-0.4426148	0.323131	-0.3053408	-0.7543519	-0.3297359	-0.5188051	-0.0261177	0.1787795
ITEM108C	-0.0670796	-0.1159786	-0.2359265	-0.2725637	0.1226551	-0.0881767	0.1575084	-0.2431494	0.0144318	-0.0088968
ITEM2	0.4042509	0.1753217	0.1270368	-0.1271776	-0.1624466	-0.0954278	0.3237726	-0.0117736	-0.1145417	-0.1420204
ITEM27	0.3418298	-0.6614568	0.6504821	-0.2772775	0.0680847	0.5170462	0.3035638	0.5205166	-0.1145838	-0.2801802
ITEM3	0.0589588	0.0728568	0.0209989	-0.0309104	-0.0868817	-0.077589	-0.1386794	0.1293244	0.0226631	-0.134611
ITEM31	0.2284335	-0.903605	0.5391906	-0.4323478	0.2203958	0.7698994	0.4315045	0.6504145	-0.0677645	-0.304926
ITEM36A	0.2278444	-0.7724653	0.4944016	-0.26389	0.174518	0.622784	0.2820259	0.5788049	-0.0847161	-0.2404074
ITEM36D	0.2284383	-0.709998	0.4291178	-0.3231784	0.161539	0.6296259	0.3626731	0.5516007	-0.1286574	-0.2464428
ITEM37	1	-0.0809245	0.2300319	-0.2335768	0.0873791	0.1057689	0.2328068	0.1128164	0.0704058	-0.0947348
ITEM41	-0.0809245	1	-0.3445638	0.3312688	-0.2065031	-0.6499944	-0.3190144	-0.6646492	0.0743993	0.3582519
ITEM43A	0.2300319	-0.3445638	1	-0.4742844	-0.0287079	0.5002545	0.3790226	0.3934731	-0.0610627	-0.2536528
ITEM43B	-0.2335768	0.3312688	-0.4742844	1	-0.1093647	-0.3360834	-0.5649335	-0.182378	-0.0108744	0.3352326
ITEM45	0.0873791	-0.2065031	-0.0287079	-0.1093647	1	0.2043639	-0.0379427	0.0221447	0.3547725	0.0481721
ITEM48	0.1057689	-0.6499944	0.5002545	-0.3360834	0.2043639	1	0.4847782	0.4628996	0.0193143	-0.1392129
ITEM55B	0.2328068	-0.3190144	0.3790227	-0.5649335	-0.0379427	0.4847782	1	0.213724	-0.3432153	-0.2576791
ITEM91	0.1128164	-0.6646492	0.3934732	-0.182378	0.0221447	0.4628996	0.213724	1	-0.0645724	-0.4058793
ITEM92B	0.0704058	0.0743993	-0.0610627	-0.0108744	0.3547725	0.0193143	-0.3432153	-0.0645724	1	0.159102
ITEM92C	-0.0947348	0.3582519	-0.2536528	0.3352326	0.0481721	-0.1392129	-0.2576791	-0.4058793	0.159102	1

Region 5

	ITEM106	ITEM107	ITEM108A	ITEM108C	ITEM2	ITEM27	ITEM3	ITEM31	ITEM36A	ITEM36D
ITEM106	1	-0.2059787	-0.1442554	-0.1654342	-0.140062	-0.4362526	-0.0368625	0.2020554	0.1135648	0.2754389
ITEM107	-0.2059787	1	0.7416897	-0.0245017	0.1301815	-0.4849357	0.0027025	-0.695103	-0.5917969	-0.5260467
ITEM108A	-0.1442554	0.7416897	1	-0.2496549	0.1879214	-0.6418702	0.1498732	-0.7447982	-0.7104894	-0.5545425
ITEM108C	-0.1654342	-0.0245017	-0.2496549	1	-0.1376859	0.3216078	-0.1807967	0.2120691	0.239281	0.1675658
ITEM2	-0.140062	0.1301815	0.1879214	-0.1376859	1	-0.0639688	0.1906971	-0.1310147	-0.0287601	-0.0257
ITEM27	-0.4362526	-0.4849357	-0.6418702	0.3216078	-0.0639688	1	-0.0294819	0.7122228	0.6944957	0.495017
ITEM3	-0.0368625	0.0027025	0.1498732	-0.1807967	0.1906971	-0.0294819	1	-0.0404667	-0.0452142	0.0093359
ITEM31	0.2020554	-0.695103	-0.7447982	0.2120691	-0.1310147	0.7122228	-0.0404667	1	0.7825162	0.7057811
ITEM36A	0.1135648	-0.5917969	-0.7104894	0.239281	-0.0287601	0.6944957	-0.0452142	0.7825162	1	0.8364081
ITEM36D	0.2754389	-0.5260467	-0.5545425	0.1675658	-0.0257	0.495017	0.0093359	0.7057811	0.8364081	1
ITEM37	-0.2255352	-0.4409292	-0.418134	0.1133946	-0.0478935	0.646754	-0.0705598	0.4938191	0.4193814	0.2650318
ITEM41	-0.0871699	0.6405325	0.6751856	-0.2344672	0.1248841	-0.7362316	0.0599275	-0.8647668	-0.7330004	-0.6412757
ITEM43A	-0.0847985	0.0360246	-0.1483828	0.0134136	-0.0893379	0.4128287	0.0015013	0.3364837	0.2197024	0.1537619
ITEM43B	-0.1786258	0.5208402	0.5944662	-0.2006361	0.191967	-0.431046	0.1356931	-0.5970764	-0.4821637	-0.4027285
ITEM45	0.4625249	-0.4559538	-0.4136169	-0.0104956	-0.2097405	0.120758	-0.0657959	0.5068213	0.350125	0.4611781
ITEM48	0.3913522	-0.3413804	-0.3728689	0.0269264	-0.1772737	0.0790649	-0.0652293	0.4180541	0.2821109	0.3767232
ITEM55B	0.2640386	-0.1825813	-0.1351945	0.0015952	-0.3955241	-0.0172322	-0.1088643	0.1677625	-0.0562593	0.0339302
ITEM91	-0.3574145	-0.1879965	-0.1966442	-0.0086095	0.228195	0.560945	0.1833364	0.2839684	0.4424252	0.3489271
ITEM92B	0.2128681	-0.075674	-0.0793547	-0.0457744	-0.0533906	-0.1622014	-0.0537259	0.0202427	-0.0059089	0.0323372
ITEM92C	0.3980752	-0.1329901	-0.1374537	-0.092048	-0.127276	-0.1343343	-0.0409452	0.1520537	0.0255611	0.1286491

Region 5 - Continued

	ITEM37	ITEM41	ITEM43A	ITEM43B	ITEM45	ITEM48	ITEM55B	ITEM91	ITEM92B	ITEM92C
ITEM106	-0.2255352	-0.0871699	-0.0847985	-0.1786258	0.4625249	0.3913522	0.2640386	-3.57E-01	0.2128681	0.3980752
ITEM107	-0.4409292	0.6405325	0.0360246	0.5208402	-0.4559538	-0.3413804	-0.1825813	-1.88E-01	-0.075674	-0.1329901
ITEM108A	-0.418134	0.6751856	-0.1483828	0.5944662	-0.4136169	-0.3728689	-0.1351945	-1.97E-01	-0.0793547	-0.1374537
ITEM108C	0.1133946	-0.2344672	0.0134136	-0.2006361	-0.0104956	0.0269264	0.0015952	-8.61E-03	-0.0457744	-0.092048
ITEM2	-0.0478935	0.1248841	-0.0893379	0.191967	-0.2097405	-0.1772737	-0.3955241	2.28E-01	-0.0533906	-0.127276
ITEM27	0.646754	-0.7362316	0.4128287	-0.431046	0.120758	0.0790649	-0.0172322	5.61E-01	-0.1622014	-0.1343343
ITEM3	-0.0705598	0.0599275	0.0015013	0.1356931	-0.0657959	-0.0652293	-0.1088643	1.83E-01	-0.0537259	-0.0409452
ITEM31	0.4938191	-0.8647668	0.3364837	-0.5970764	0.5068213	0.4180541	0.1677625	2.84E-01	0.0202427	0.1520537
ITEM36A	0.4193814	-0.7330004	0.2197024	-0.4821637	0.350125	0.2821109	-0.0562593	4.42E-01	-0.0059089	0.0255611
ITEM36D	0.2650318	-0.6412757	0.1537619	-0.4027285	0.4611781	0.3767232	0.0339302	3.49E-01	0.0323372	0.1286491
ITEM37	1	-0.5648511	0.2288802	-0.529333	0.0804049	-0.1945472	0.0203116	3.27E-01	-0.3077851	-0.1755586
ITEM41	-0.5648511	1	-0.2966947	0.5201109	-0.4043646	-0.2880242	-0.0532497	-3.60E-01	-0.0004099	-0.0387844
ITEM43A	0.2288802	-0.2966947	1	-0.2047478	0.1843203	0.2312268	0.2555333	2.20E-01	-0.0325376	0.0598795
ITEM43B	-0.529333	0.5201109	-0.2047478	1	-0.4169554	-0.1786511	-0.2613467	2.91E-02	0.1164397	-0.1194337
ITEM45	0.0804049	-0.4043646	0.1843203	-0.4169554	1	0.636317	0.402788	-1.15E-01	0.2909067	0.3685513
ITEM48	-0.1945472	-0.2880242	0.2312268	-0.1786511	0.636317	1	0.370542	-1.62E-01	0.5178746	0.4671781
ITEM55B	0.0203116	-0.0532497	0.2555333	-0.2613467	0.402788	0.370542	1	-3.81E-01	-0.1424508	0.3491075
ITEM91	0.3268732	-0.3598606	0.2198844	0.0291029	-0.1146799	-0.1618401	-0.3811808	1.00E+00	-0.1455354	-0.2335899
ITEM92B	-0.3077851	-0.0004099	-0.0325376	0.1164397	0.2909067	0.5178746	-0.1424508	-1.46E-01	1	0.1873888
ITEM92C	-0.1755586	-0.0387844	0.0598795	-0.1194337	0.3685513	0.4671781	0.3491075	-2.34E-01	0.1873888	1

Region 9

	ITEM106	ITEM107	ITEM108A	ITEM108C	ITEM2	ITEM27	ITEM3	ITEM31	ITEM36A	ITEM36D
ITEM106	1	0.0515731	0.1991034	-0.0912928	-0.1487494	-0.3704439	0.2186844	-0.1821271	-0.1260626	-0.1255828
ITEM107	0.0515731	1	0.7693032	0.5908317	-0.500306	-0.6811979	0.468825	-0.8162641	-0.7074456	-0.6426781
ITEM108A	0.1991034	0.7693032	1	0.5012122	-0.3374362	-0.7470653	0.3904683	-0.7609821	-0.6401213	-0.5344412
ITEM108C	-0.0912928	0.5908317	0.5012122	1	-0.3580977	-0.3595257	0.1599243	-0.4082289	-0.3500412	-0.2916991
ITEM2	-0.1487494	-0.500306	-0.3374362	-0.3580977	1	0.6294943	-0.8160046	0.6375079	0.717339	0.7558247
ITEM27	-0.3704439	-0.6811979	-0.7470653	-0.3595257	0.6294943	1	-0.6880628	0.9125598	0.8768207	0.7779948
ITEM3	0.2186844	0.468825	0.3904683	0.1599243	-0.8160046	-0.6880628	1	-0.6514062	-0.7055584	-0.7732569
ITEM31	-0.182127	-0.8162641	-0.7609821	-0.4082289	0.6375079	0.9125598	-0.6514062	1	0.9006072	0.8199397
ITEM36A	-0.1260626	-0.7074455	-0.6401213	-0.3500412	0.717339	0.8768207	-0.7055584	0.9006072	1	0.897421
ITEM36D	-0.1255828	-0.6426781	-0.5344412	-0.2916991	0.7558247	0.7779948	-0.7732569	0.8199397	0.897421	1
ITEM37	-0.1074087	-0.5304335	-0.4294321	-0.3444081	0.3101907	0.5894805	-0.3085607	0.545126	0.4873166	0.3907729
ITEM41	0.1164189	0.7402679	0.6074233	0.352041	-0.4056428	-0.6603489	0.3439957	-0.7792045	-0.6651046	-0.5510133
ITEM43A	-0.2645228	-0.1075848	-0.3351178	-0.1933225	0.4874322	0.6214988	-0.472089	0.3990726	0.471288	0.4230073
ITEM43B	-0.0425631	0.1691834	-0.0336372	0.1063252	-0.1880186	-0.0962666	0.144225	-0.0322014	-0.0490327	-0.0501572
ITEM45	0.0527473	-0.2743885	-0.3071847	-0.2011235	0.0650394	0.1407845	-0.0453793	0.2228616	0.1599428	0.191155
ITEM48	-0.1222298	-0.392125	-0.5269151	-0.2745213	0.2252033	0.4453907	-0.2491733	0.4864784	0.4231626	0.4048457
ITEM55B	-0.1597306	-0.7318557	-0.7096728	-0.4706204	0.6767141	0.8320764	-0.6135979	0.9084997	0.8537759	0.7643998
ITEM91	-0.1904265	-0.3813255	-0.3248004	-0.1448172	-0.0403779	0.254251	0.037267	0.3378399	0.2242462	0.0962109
ITEM92B	0.1063716	0.1196251	0.1002234	0.0142027	-0.2100921	-0.2875454	0.1790391	-0.2420518	-0.2772794	-0.1905533
ITEM92C	0.0801078	0.0063781	0.051732	-0.0596044	-0.0116884	-0.0805171	-0.0010606	-0.0486956	-0.0607545	-0.0229749

Region 9 - Continued

	ITEM37	ITEM41	ITEM43A	ITEM43B	ITEM45	ITEM48	ITEM55B	ITEM91	ITEM92B	ITEM92C
ITEM106	-0.1074087	0.1164189	-0.2645228	-0.0425631	0.0527473	-0.1222298	-0.1597306	-0.1904265	0.1063716	0.0801078
ITEM107	-0.5304335	0.7402679	-0.1075848	0.1691834	-0.2743885	-0.392125	-0.7318557	-0.3813255	0.1196251	0.0063781
ITEM108A	-0.4294321	0.6074233	-0.3351178	-0.0336372	-0.3071847	-0.5269151	-0.7096728	-0.3248004	0.1002234	0.051732
ITEM108C	-0.3444081	0.352041	-0.1933225	0.1063252	-0.2011235	-0.2745213	-0.4706204	-0.1448172	0.0142027	-0.0596044
ITEM2	0.3101907	-0.4056428	0.4874322	-0.1880186	0.0650394	0.2252033	0.6767141	-0.0403779	-0.2100921	-0.0116884
ITEM27	0.5894805	-0.6603489	0.6214988	-0.0962666	0.1407845	0.4453907	0.8320764	0.254251	-0.2875454	-0.0805171
ITEM3	-0.3085607	0.3439957	-0.472089	0.144225	-0.0453793	-0.2491733	-0.6135979	0.037267	0.1790391	-0.0010606
ITEM31	0.545126	-0.7792045	0.3990726	-0.0322014	0.2228616	0.4864784	0.9084997	0.3378399	-0.2420518	-0.0486956
ITEM36A	0.4873166	-0.6651046	0.471288	-0.0490327	0.1599428	0.4231626	0.8537759	0.2242462	-0.2772794	-0.0607545
ITEM36D	0.3907729	-0.5510133	0.4230073	-0.0501572	0.191155	0.4048457	0.7643998	0.0962109	-0.1905533	-0.0229749
ITEM37	1	-0.5738168	0.1855316	-0.5529143	0.0317036	-0.1059785	0.4289536	0.2743777	-0.3292927	-0.0002456
ITEM41	-0.5738168	1	-0.1365536	0.115058	-0.2044957	-0.3241669	-0.674044	-0.516158	0.1788208	0.0291306
ITEM43A	0.1855316	-0.1365536	1	-0.0030402	0.0467586	0.3490447	0.4359049	-0.1896207	-0.116914	-0.0307394
ITEM43B	-0.5529143	0.115058	-0.0030402	1	0.0508983	0.5694654	0.0977655	-0.0224091	0.2492295	-0.1388423
ITEM45	0.0317036	-0.2044957	0.0467586	0.0508983	1	0.3695074	0.2123365	0.0139207	0.204675	0.0224329
ITEM48	-0.1059785	-0.3241669	0.3490447	0.5694654	0.3695074	1	0.5185668	0.0631305	0.2522024	-0.0482679
ITEM55B	0.4289536	-0.674044	0.4359049	0.0977655	0.2123365	0.5185668	1	0.278648	-0.3234148	-0.0654699
ITEM91	0.2743777	-0.516158	-0.1896207	-0.0224091	0.0139207	0.0631305	0.278648	1	-0.1632333	-0.1445839
ITEM92B	-0.3292927	0.1788208	-0.116914	0.2492295	0.204675	0.2522024	-0.3234148	-0.1632333	1	0.0488434
ITEM92C	-0.0002456	0.0291306	-0.0307394	-0.1388423	0.0224329	-0.0482679	-0.0654699	-0.1445839	0.0488434	1

Region 10

	ITEM106	ITEM107	ITEM108A	ITEM108C	ITEM2	ITEM27	ITEM3	ITEM31	ITEM36A	ITEM36D
ITEM106	1	-0.2557984	0.0292817	-0.1573376	-0.1826241	-0.5759988	-0.008119	-0.3009488	0.0265755	0.0792229
ITEM107	-0.2557984	1	0.6949758	0.0906834	0.099673	0.0807616	0.2128324	-0.3105187	-0.5477722	-0.5571526
ITEM108A	0.0292817	0.6949758	1	-0.0418902	0.175904	-0.4141278	0.0735153	-0.6069072	-0.6526613	-0.616984
ITEM108C	-0.1573376	0.0906834	-0.0418902	1	0.0211728	0.322884	0.0540707	0.2570785	0.2719572	0.1888099
ITEM2	-0.1826241	0.099673	0.1759041	0.0211728	1	0.0856967	-0.3258199	0.0995318	-0.2213943	-0.2793783
ITEM27	-0.5759988	0.0807616	-0.4141278	0.322884	0.0856967	1	-0.0710279	0.8206921	0.482509	0.3480095
ITEM3	-0.008119	0.2128324	0.0735153	0.0540707	-0.3258199	-0.0710279	1	-0.1388663	-0.1208863	-0.1470643
ITEM31	-0.3009488	-0.3105187	-0.6069072	0.2570785	0.0995318	0.8206921	-0.1388663	1	0.6443341	0.4954949
ITEM36A	0.0265755	-0.5477722	-0.6526613	0.2719572	-0.2213943	0.482509	-0.1208863	0.6443341	1	0.877805
ITEM36D	0.0792229	-0.5571526	-0.616984	0.1888099	-0.2793783	0.3480095	-0.1470643	0.4954949	0.877805	1
ITEM37	-0.2634975	-0.0860457	-0.2404514	0.1469137	0.0975937	0.5576997	0.0006372	0.5626435	0.3787837	0.2729634
ITEM41	0.1859852	0.3880345	0.5008812	-0.2074945	-0.0183201	-0.5455063	0.147025	-0.7177788	-0.5836837	-0.4623119
ITEM43A	-0.3590668	0.4763034	0.0300453	0.0756391	0.1529773	0.602237	-0.0315379	0.3329637	-0.0294292	-0.1047575
ITEM43B	0.2590382	-0.0122364	0.0554644	-0.0278518	-0.1655687	-0.4534341	0.1141213	-0.3687031	-0.1301993	-0.0509163
ITEM45	0.2446958	-0.4616851	-0.3330694	-0.1229285	-0.1641252	-0.1856946	-0.0185284	0.0267862	0.2212431	0.2938428
ITEM48	0.1388905	-0.5802866	-0.6143224	0.0109008	-0.1588053	0.0210299	-0.1064117	0.2696025	0.4090578	0.4236652
ITEM55B	0.169624	-0.7096172	-0.6738236	0.0225321	-0.1565565	0.1450166	0.0241494	0.4325527	0.4986431	0.505533
ITEM91	-0.1043605	-0.4102441	-0.4803862	0.127649	-0.1418369	0.3988392	-0.0371512	0.5635977	0.5370455	0.4556137
ITEM92B	0.1458963	-0.0610606	0.0295905	-0.0893012	-0.1163339	-0.3175968	0.0116004	-0.2402459	-0.072871	-0.0409541
ITEM92C	0.1847325	-0.2138433	-0.1385375	-0.0458957	-0.2278279	-0.1843637	-0.0370617	-0.0673057	0.1663628	0.2147853

Region 10 - Continued

	ITEM37	ITEM41	ITEM43A	ITEM43B	ITEM45	ITEM48	ITEM55B	ITEM91	ITEM92B	ITEM92C
ITEM106	-0.2634975	0.1859852	-0.3590668	0.2590382	0.2446958	0.1388905	0.169624	-0.1043605	0.1458963	0.1847325
ITEM107	-0.0860457	0.3880345	0.4763034	-0.0122364	-0.4616851	-0.5802866	-0.7096172	-0.4102441	-0.0610606	-0.2138433
ITEM108A	-0.2404514	0.5008812	0.0300453	0.0554644	-0.3330694	-0.6143224	-0.6738236	-0.4803862	0.0295905	-0.1385375
ITEM108C	0.1469137	-0.2074945	0.0756391	-0.0278518	-0.1229285	0.0109008	0.0225321	0.127649	-0.0893012	-0.0458957
ITEM2	0.0975937	-0.0183201	0.1529773	-0.1655687	-0.1641252	-0.1588053	-0.1565565	-0.1418369	-0.1163339	-0.2278279
ITEM27	0.5576997	-0.5455063	0.602237	-0.4534341	-0.1856946	0.0210299	0.1450166	0.3988392	-0.3175968	-0.1843637
ITEM3	0.0006372	0.147025	-0.0315379	0.1141213	-0.0185284	-0.1064117	0.0241494	-0.0371512	0.0116004	-0.0370617
ITEM31	0.5626435	-0.7177788	0.3329637	-0.3687031	0.0267862	0.2696025	0.4325527	0.5635977	-0.2402459	-0.0673057
ITEM36A	0.3787837	-0.5836837	-0.0294292	-0.1301993	0.2212431	0.4090578	0.4986431	0.5370455	-0.072871	0.1663628
ITEM36D	0.2729634	-0.4623119	-0.1047575	-0.0509163	0.2938428	0.4236652	0.505533	0.4556137	-0.0409541	0.2147853
ITEM37	1	-0.5298411	0.3003805	-0.5738871	0.0345564	-0.2237606	0.3246718	0.5177429	-0.4119498	-0.3080941
ITEM41	-0.5298411	1	-0.0538194	0.2841073	-0.0700294	-0.2162185	-0.3815852	-0.6702496	0.1742586	0.0155512
ITEM43A	0.3003805	-0.0538194	1	-0.4591355	-0.2046904	-0.1488229	-0.0837247	0.0155206	-0.1940236	-0.2123317
ITEM43B	-0.5738871	0.2841073	-0.4591355	1	-0.0036144	0.3794375	-0.0613148	-0.2620181	0.3435012	0.4099339
ITEM45	0.0345564	-0.0700294	-0.2046904	-0.0036144	1	0.3172055	0.4196998	0.1351657	0.2513561	0.1791125
ITEM48	-0.2237606	-0.2162185	-0.1488229	0.3794375	0.3172055	1	0.5325245	0.1867824	0.3975538	0.5392108
ITEM55B	0.3246718	-0.3815852	-0.0837247	-0.0613148	0.4196998	0.5325246	1	0.4929841	-0.1654117	0.0998441
ITEM91	0.5177429	-0.6702496	0.0155206	-0.2620181	0.1351657	0.1867824	0.4929841	1	-0.2540023	-0.0561407
ITEM92B	-0.4119498	0.1742586	-0.1940236	0.3435012	0.2513561	0.3975538	-0.1654117	-0.2540023	1	0.5450289
ITEM92C	-0.3080941	0.0155512	-0.2123317	0.4099339	0.1791125	0.5392108	0.0998441	-0.0561407	0.5450289	1

Appendix E

Case II Error-Model Regression Output

Summary Regression Output for All Data

Regression Statistics	
Multiple R	0.703038
R Square	0.494263
Adjusted R Square	0.493274
Standard Error	0.497924
Observations	2052

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	495.9944	123.9986	500.1390	4.7003E-301
Residual	2047	507.5092	0.2479		
Total	2051	1003.5035			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.44013	0.04591	9.58742	2.51429E-21
Distance	0.00635	0.00036	17.82581	3.4593E-66
Resolution	-0.18785	0.01063	-17.66614	4.06296E-65
Color	0.12147	0.01649	7.36719	2.51412E-13
Intensity	-0.00304	0.00040	-7.69441	2.19423E-14

4-Fold Summary Regression Output for All Colors 1/4

Regression Statistics	
Multiple R	0.702383
R Square	0.493341
Adjusted R Square	0.492020
Standard Error	0.499100
Observations	1539

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	372.0774	93.0193	373.4197	1.2471E-224
Residual	1534	382.1214	0.2491		
Total	1538	754.1988			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.44577	0.05254	8.48363	5.06345E-17
Distance	0.00623	0.00041	15.15537	1.8742E-48
Resolution	-0.18556	0.01233	-15.05245	7.29621E-48
Color	0.12922	0.01897	6.81258	1.37249E-11
Intensity	-0.00318	0.00046	-6.96296	4.9243E-12

4-Fold Summary Regression Output for All Colors 2/4

Regression Statistics	
Multiple R	0.700551758
R Square	0.490772765
Adjusted R Square	0.489444058
Standard Error	0.500353707
Observations	1538

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	369.8837304	92.47093259	369.3609639	8.4005E-223
Residual	1533	383.7924239	0.250353832		
Total	1537	753.6761543			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.511049304	0.053620825	9.530798951	5.86213E-21
Distance	0.005896897	0.000417114	14.13737499	9.56738E-43
Resolution	-0.192054272	0.012290585	-15.62612921	3.45183E-51
Color	0.104337275	0.019384176	5.38260045	8.48243E-08
Intensity	-0.003397446	0.000457866	-7.420178093	1.92906E-13

4-Fold Summary Regression Output for All Colors 3/4

Regression Statistics	
Multiple R	0.687857493
R Square	0.47314793
Adjusted R Square	0.468999489
Standard Error	0.504785782
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	116.2481886	29.06204716	114.0543818	2.46807E-69
Residual	508	129.4428125	0.254808686		
Total	512	245.6910012			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	0.438010845	0.095141748	4.603771244	5.24921E-06	0.251091107
Distance	0.006216151	0.000731212	8.50116716	2.10233E-16	0.00477958
Resolution	-0.179354665	0.021038728	-8.524976765	1.75643E-16	-0.220688292
Color	0.108915786	0.034224541	3.182388491	0.001550079	0.04167672
Intensity	-0.003053502	0.000799773	-3.817962111	0.000151176	-0.004624772

4-Fold Summary Regression Output for All Colors 4/4

Regression Statistics	
Multiple R	0.707962503
R Square	0.501210906
Adjusted R Square	0.499910282
Standard Error	0.49629013
Observations	1539

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	379.6646825	94.91617064	385.362039	7.7312E-230
Residual	1534	377.8301727	0.246303894		
Total	1538	757.4948553			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	0.440989037	0.052474562	8.40386306	9.72252E-17	0.338059572
Distance	0.006385598	0.000408065	15.64848318	2.5369E-51	0.005585173
Resolution	-0.190948585	0.01235183	-15.45913378	3.2631E-50	-0.215176843
Color	0.126061443	0.018848413	6.688172957	3.15622E-11	0.089090062
Intensity	-0.003035691	0.000456142	-6.655140576	3.92808E-11	-0.00393042

4-Fold Summary Regression Output for White 1/4

Regression Statistics	
Multiple R	0.761890704
R Square	0.580477446
Adjusted R Square	0.577174118
Standard Error	0.282787532
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	56.21008755	14.05252189	175.7250827	2.24597E-94
Residual	508	40.62414432	0.079968788		
Total	512	96.83423187			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.731225496	0.049402653	14.80134081	1.79839E-41
Distance	0.000651392	0.000473491	1.375722711	0.169513745
Resolution	-0.118101559	0.012570927	-9.39481729	1.93685E-19
Color	0	0	65535	#NUM!
Intensity	-0.003974937	0.000370912	-10.7166689	#NUM!

4-Fold Summary Regression Output for White 2/4

Regression Statistics	
Multiple R	0.759249305
R Square	0.576459507
Adjusted R Square	0.573124543
Standard Error	0.281927361
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	54.95561016	13.73890254	172.8532659	2.51115E-93
Residual	508	40.37738283	0.079483037		
Total	512	95.33299299			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.781755162	0.047999488	16.28673967	2.71958E-48
Distance	0.000177774	0.000476426	0.373140328	0.709199654
Resolution	-0.121928916	0.012531737	-9.72961014	1.24828E-20
Color	0	0	65535	#NUM!
Intensity	-0.004178254	0.000367839	-11.3589353	#NUM!

4-Fold Summary Regression Output for White 3/4

Regression Statistics	
Multiple R	0.773473316
R Square	0.598260971
Adjusted R Square	0.595097671
Standard Error	0.27426577
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	56.90541231	14.22635308	189.1256207	3.8579E-99
Residual	508	38.21262999	0.075221713		
Total	512	95.11804229			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.770307592	0.047504674	16.21540618	5.84775E-48
Distance	0.000334539	0.000462762	0.722919056	0.47006231
Resolution	-0.118670568	0.011891982	-9.97904003	1.5529E-21
Color	0	0	65535	#NUM!
Intensity	-0.004215041	0.000357482	-11.7909271	#NUM!

4-Fold Summary Regression Output for White 4/4

Regression Statistics	
Multiple R	0.758943521
R Square	0.575995268
Adjusted R Square	0.572656649
Standard Error	0.280868905
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	54.44014176	13.61003544	172.5249594	3.31413E-93
Residual	508	40.07476963	0.078887342		
Total	512	94.5149114			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Multiple R	0.775058538	0.047236773	16.40794848	7.38798E-49
R Square	0.000187504	0.000463104	0.40488566	0.685731994
Adjusted R Square	-0.125598938	0.012576955	-9.9864344	1.45908E-21
Standard Error	0	0	65535	#NUM!
Observations	-0.004092744	0.000360018	-11.3681733	#NUM!

4-Fold Summary Regression Output for Gray 1/4

Regression Statistics	
Multiple R	0.843063189
R Square	0.710755541
Adjusted R Square	0.708478026
Standard Error	0.400698809
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	200.4264442	50.10661106	312.0749631	2.6319E-135
Residual	508	81.56424393	0.160559535		
Total	512	281.9906882			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.140754692	0.065348468	17.4564872	8.3863E-54
Distance	0.003277323	0.000639659	5.123545337	4.26645E-07
Resolution	-0.208252632	0.018562004	-11.21929666	3.00148E-26
Color	0	0	65535	#NUM!
Intensity	-0.010911317	0.00086786	-12.57267223	#NUM!

4-Fold Summary Regression Output for Gray 2/4

Regression Statistics	
Multiple R	0.83462912
R Square	0.696605767
Adjusted R Square	0.694216836
Standard Error	0.409243939
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	195.3475502	48.83688755	291.5972782	4.7877E-130
Residual	508	85.08014555	0.167480601		
Total	512	280.4276958			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.131517589	0.065996198	17.1451936	2.51296E-52
Distance	0.003214604	0.000646273	4.974063711	8.98587E-07
Resolution	-0.191255819	0.018398859	-10.39498259	4.4543E-23
Color	0	0	65535	#NUM!
Intensity	-0.0109859	0.000873798	-12.57258332	#NUM!

4-Fold Summary Regression Output for Gray 3/4

Regression Statistics	
Multiple R	0.836809888
R Square	0.700250788
Adjusted R Square	0.697890558
Standard Error	0.390319206
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	180.8002858	45.20007144	296.687519	2.2339E-131
Residual	508	77.39333413	0.152349083		
Total	512	258.1936199			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.139743226	0.066550916	17.12588345	3.10141E-52
Distance	0.003004677	0.000639632	4.697507426	3.39447E-06
Resolution	-0.206574083	0.017875996	-11.55594838	1.38351E-27
Color	0	0	65535	#NUM!
Intensity	-0.010700253	0.000865297	-12.36598582	#NUM!

4-Fold Summary Regression Output for Gray 4/4

Regression Statistics	
Multiple R	0.829032888
R Square	0.687295529
Adjusted R Square	0.684828433
Standard Error	0.413827572
Observations	512

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	190.8342112	47.7085528	278.5847866	1.8201E-126
Residual	507	86.82540265	0.17125326		
Total	511	277.6596139			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.186738629	0.067747701	17.51703175	4.53567E-54
Distance	0.002738737	0.000667437	4.103366056	4.74395E-05
Resolution	-0.216213547	0.019421408	-11.13274332	6.63877E-26
Color	0	0	65535	#NUM!
Intensity	-0.011120144	0.000896958	-12.3976244	#NUM!

4-Fold Summary Regression Output for Black 1/4

Regression Statistics	
Multiple R	0.694769883
R Square	0.482705191
Adjusted R Square	0.478632003
Standard Error	0.584284862
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	161.8291786	40.45729466	118.5079729	2.40728E-71
Residual	508	173.4255104	0.3413888		
Total	512	335.2546891			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.974943814	0.101398499	9.614972834	3.21521E-20
Distance	0.004081946	0.001018599	4.007410734	7.05809E-05
Resolution	-0.074410075	0.02494122	-2.983417625	0.002987584
Color	0	0	65535	#NUM!
Intensity	-0.018020796	0.002276631	-7.915555238	#NUM!

4-Fold Summary Regression Output for Black 2/4

Regression Statistics	
Multiple R	0.699918988
R Square	0.48988659
Adjusted R Square	0.485869949
Standard Error	0.585397445
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	167.1837913	41.79594783	121.9642452	7.01064E-73
Residual	508	174.0866059	0.342690169		
Total	512	341.2703973			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.997182551	0.103932608	9.594511001	3.80371E-20
Distance	0.003802901	0.001027626	3.700666588	0.000238522
Resolution	-0.059557608	0.025942329	-2.29576954	0.022095486
Color	0	0	65535	#NUM!
Intensity	-0.018933474	0.002357338	-8.031717168	#NUM!

4-Fold Summary Regression Output for Black 3/4

Regression Statistics	
Multiple R	0.711061442
R Square	0.505608374
Adjusted R Square	0.501715527
Standard Error	0.579207466
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	174.2910822	43.57277054	129.8813736	2.54729E-76
Residual	508	170.4244944	0.335481288		
Total	512	344.7155766			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.859545101	0.10418191	8.250425593	1.36513E-15
Distance	0.005421312	0.001015594	5.338072242	1.41856E-07
Resolution	-0.074101321	0.025034272	-2.959995014	0.003220145
Color	0	0	65535	#NUM!
Intensity	-0.015973714	0.002322162	-6.878809904	#NUM!

4-Fold Summary Regression Output for Black 4/4

Regression Statistics	
Multiple R	0.691386676
R Square	0.478015535
Adjusted R Square	0.473905421
Standard Error	0.598343682
Observations	513

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	166.5518898	41.63797245	116.30226	2.35973E-70
Residual	508	181.8717022	0.358015162		
Total	512	348.4235921			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.065799433	0.103045444	10.3430039	6.97785E-23
Distance	0.003136527	0.001013583	3.094493855	0.002080142
Resolution	-0.064558445	0.025613043	-2.520530125	0.012023032
Color	0	0	65535	#NUM!
Intensity	-0.020251612	0.002302733	-8.794600249	#NUM!