Forensic Application of Chemometric Analysis to Visible Absorption Spectra Collected from Dyed Textile Fibers

Alejandra Flores

University of Central Florida

Part of the Chemistry Commons

Find similar works at: https://stars.library.ucf.edu/etd

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

https://stars.library.ucf.edu/etd/1343
FORENSIC APPLICATION OF CHEMOMETRIC ANALYSIS TO VISIBLE ABSORPTION SPECTRA COLLECTED FROM DYED TEXTILE FIBERS

by

ALEJANDRA FLORES
B.S. Indiana University-Purdue University Indianapolis, 2011
B.A. Indiana University-Purdue University Indianapolis, 2011
M.S. University of Central Florida, 2013

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the College of Sciences at the University of Central Florida Orlando, Florida

Spring Term
2015

Major Professor: Michael E. Sigman
ABSTRACT

Forensic analysis of evidence consists of the comparison of physical, spectroscopic, or chemical characteristics of a questioned sample to a set of knowns. Currently, decisions as to whether or not the questioned sample can be associated or grouped with the knowns are left up to the discretion of the forensic analyst. The implications of these outcomes are presented as evidence to a jury in a court of law to determine if a defendant is guilty of committing a crime or not. Leading up to, and since, the publication of the National Academy of Sciences (NAS) report entitled “Strengthening Forensic Science in the United States: A Path Forward,” the inadequacies of allowing potentially biased forensic opinion to carry such weight in the courtroom have been unmasked. This report exposed numerous shortcomings in many areas of forensic science, but also made recommendations on how to fortify the discipline. The main suggestions directed towards disciplines that analyze trace evidence include developing error rates for commonly employed practices and evaluating method reliability and validity.

This research focuses on developing a statistical method of analysis for comparing visible absorption profiles collected from highly similarly colored textile fibers via microspectrophotometry (MSP). Several chemometric techniques were applied to spectral data and utilized to help discriminate fibers beyond the point where traditional methods of microscopical examination may fail. Because a dye’s chemical structure dictates the shape of the absorption profile, two fibers dyed with chemically similar dyes
can be very difficult to distinguish from one another using traditional fiber examination techniques. The application of chemometrics to multivariate spectral data may help elicit latent characteristics that may aid in fiber discrimination.

The three sample sets analyzed include dyed fabric swatches (three pairs of fabrics were dyed with chemically similar dye pairs), commercially available blue yarns (100% acrylic), and denims fabrics (100% cotton). Custom dyed swatches were each dyed uniformly with a single dye whereas the dye formulation for both the yarns and denims is unknown. As a point for study, spectral comparisons were performed according to the guidelines published by the Standard Working Group for Materials Analysis (SWGMAT) Fiber Subgroup based on visual analysis only. In the next set of tests, principal components analysis (PCA) was utilized to reduce the dimensionality of the large multivariate data sets and to visualize the natural groupings of samples. Comparisons were performed using the resulting PCA scores where group membership of the questioned object was evaluated against the known objects using the score value as the distance metric. Score value is calculated using the score and orthogonal distances, the respective cutoff values based on a quantile percentage, and an optimization parameter, $\gamma$. Lastly, likelihood ratios (LR) were generated from density functions modelled from similarity values assessing comparisons between sample population data. R code was written in-house to execute all method of fiber comparisons described here. The SWGMAT method performed with 62.7% accuracy, the optimal accuracy rate for the score value method was 75.9%, and the accuracy
rates for swatch-yarn and denim comparisons, respectively, are 97.7% and 67.1% when the LR method was applied.
This work is dedicated to Bullet.
ACKNOWLEDGMENTS

With God all things are possible.

Thank you to my advisor, Dr. Sigma, and the rest of my committee members, Dr. Campiglia, Dr. Yestrebsky, Dr. Torres, and Dr. Ni, for their help and support. Special thanks to my friends Dana-Marie, Shanna, Erin, and Jes for helping me laugh along the way.

To Jay, Dad, Mom, and Andrew- I love you.
# TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................. xii

LIST OF TABLES ............................................................................................................... xv

LIST OF EQUATIONS ....................................................................................................... xvii

CHAPTER 1: INTRODUCTION .......................................................................................... 1

1.1 Textile Fibers as Trace Evidence ........................................................................... 1

1.2 SWGMAT Guidelines ............................................................................................ 4

1.3 NAS Report .............................................................................................................. 7

1.3.1 Recommendations .......................................................................................... 8

1.4 Research Goals ....................................................................................................... 10

CHAPTER 2: BACKGROUND ......................................................................................... 13

2.1 Textile Fibers and Their Dyes ............................................................................. 13

2.1.1 Implications of Chemically Similar Dyes ..................................................... 16

2.1.2 Forensic Fiber Questions .............................................................................. 20

2.2 Fiber Analysis ........................................................................................................ 25

2.2.1 Standard Crime Laboratory Practices ......................................................... 27

2.2.2 Forensic Fiber Dye Research ...................................................................... 30

2.3 Chemometric Methods ......................................................................................... 39

2.3.1 Principal Components Analysis .................................................................. 41
2.3.2 Distance Metrics ........................................................................................................... 44
2.3.3 Likelihood Ratio ........................................................................................................... 50

CHAPTER 3: EXPERIMENTAL ................................................................................................. 54

3.1 Instrumental Parameters .................................................................................................. 54

3.2 Samples Sets .................................................................................................................... 55

3.2.1 Dyed Fabric Swatch Samples ...................................................................................... 55

3.2.2 Blue Yarn Samples ....................................................................................................... 57

3.2.3 Denim Fabric Samples ................................................................................................. 58

3.3 Sampling Methods .......................................................................................................... 58

3.3.1 Dyed Fabric Swatches ................................................................................................. 58

3.3.2 Blue Yarns .................................................................................................................... 60

3.3.3 Denim Fabrics .............................................................................................................. 62

3.4 Visual Analysis of Spectral Variance ................................................................................. 64

3.5 Visual Comparisons of Absorption Spectra ...................................................................... 64

3.6 Fiber Comparisons Using Score Value ............................................................................ 66

3.7 Fiber Comparisons Using Likelihood Ratios .................................................................... 68

CHAPTER 4: RESULTS AND DISCUSSION ............................................................................ 71

4.1 Visual Inspection of Spectral Variance .............................................................................. 71

4.1.1 Spectral Variance in Dyed Fabric Swatches ................................................................. 71
LIST OF FIGURES

Figure 2.1: Dye molecules A and B; spectral overlay of visible absorption spectra collected from solutions of A (shown in red) and B (shown in blue) .............................. 19
Figure 2.2: Red dotted line shows new latent variable of data points (black dots) after PCA is performed; green dotted lines describe variance in the orthogonal direction. Note: Adapted with permission from Introduction to Multivariate Statistical Analysis in Chemometrics (p. 61) by K. Varmuza and P. Filzmoser, Boca Raton, FL: Taylor and Francis Group, LLC. Copyright 2009(88) ....................................................................................... 42
Figure 2.3: Approximate reconstruction of $X$-matrix. Note: Adapted with permission from Introduction to Multivariate Statistical Analysis in Chemometrics (p. 62) by K. Varmuza and P. Filzmoser, Boca Raton, FL: Taylor and Francis Group, LLC. Copyright 2009(88) ................................................................................................................................................. 43
Figure 2.4: Known scores on a 2-dimensional plane in 3-dimensional space ............ 46
Figure 2.5: Questioned score$_1$ projected onto 2-dimensional plane ...................... 47
Figure 2.6: Questioned score$_2$ projected onto 2-dimensional plane ....................... 48
Figure 2.7: Coomans plot showing cutoff values for both score and orthogonal distances. Note: Adapted with permission from Introduction to Multivariate Statistical Analysis in Chemometrics (p. 212) by K. Varmuza and P. Filzmoser, Boca Raton, FL: Taylor and Francis Group, LLC. Copyright 2009(88) .................................................................................................................. 50
Figure 3.1: Molecular structures of Acid Blue 25 and 41 (top pair), Disperse Blue 3 and 14 (middle pair), and Basic Green 1 and 4 (bottom pair) .......................................................................................... 56
Figure 3.2: Generic sampling areas (1, 3, and 5) on each swatch with fabric type information written on bottom of swatch with black permanent marker ......................... 59
Figure 3.3: Generic sampling areas on denim fabrics (1, 2, 3, and 4) ....................... 62
Figure 3.4: Close up image of Den3A surface ............................................................... 63
Figure 3.5: Example of normal probability density functions ........................................ 69
Figure 3.6: Example of LR calculation from normal probability density functions ....... 70
Figure 4.1: Spectral overlays for each dyed fabric swatch, M1-M6; fiber1=red, fiber2=blue, fiber3=green, fiber4=purple, fiber5=pink, fiber6=light blue, fiber7=orange, fiber8=lilac, fiber9=gray, and fiber10=black ................................................................. 73
Figure 4.2: Spectral overlays for each blue yarn, F-J; fiber1=red, fiber2=blue, fiber3=green, fiber4=purple, fiber5=pink ........................................................................ 76
Figure 4.3: Spectral overlays for each denim fabrics, Den3A-Den15; fiber1=red, fiber2=blue, fiber3=green, fiber4=purple, fiber5=pink, fiber6=light blue ...................... 79
Figure 4.4: M1-M6 SS false spectral exclusions ............................................................. 84
Figure 4.5: F1 v F Knowns SS false spectral exclusion ............................................... 87
Figure 4.6: F1 average spectrum vs F2-F5 average spectra ......................................... 88
Figure 4.7: Den3A-Den15 SS False spectral exclusions .............................................. 90
Figure 4.8: Examples of spectral exclusion due to boundary violation ....................... 92
Figure 4.9: Example of peak inversion for M1-M6 DS comparisons ............................ 93
Figure 4.10: Two examples of false inclusion results for yarn pairs ......................... 95
Figure 4.11: Two examples of inconclusive results from yarn pairs ......................... 96
Figure 4.12: Two examples of false inclusions for DS denim comparisons ............... 99
Figure 4.13: Scores plot for M1-M6 swatches ............................................................. 102
Figure 4.14: Scores plot for F-J yarns ......................................................................... 103
Figure 4.15: Scores plot for Den3A-Den15 denims ..................................................... 105
Figure 4.16: Coomans plots for false discrimination and correct non-discrimination... 107
Figure 4.17: Blue yarns false discrimination Coomans plots ....................................... 110
Figure 4.18: pearsonr and fisher-r score density histograms ........................................ 121
Figure 4.19: pearsonr and fisher-r probability density functions .................................... 122
Figure 4.20: Swatch-yarn score distributions vs denim score distributions ................. 127
Figure 5.1: M1-M6 SS spectral comparison results.................................................... 148
Figure 5.2: F-J SS spectral comparison results............................................................ 153
Figure 5.3: Den3A-Den15 SS spectral comparison results .......................................... 163
Figure 5.4: M1-M6 DS spectral comparisons results ................................................... 177
Figure 5.5: F-J and G-I DS spectral comparisons results ............................................ 181
Figure 5.6: Den3A-Den15 DS spectral comparisons results ........................................ 189
Figure 5.7: M1-M6 SS score value results ................................................................. 194
Figure 5.8: F-J SS score value results ........................................................................ 196
Figure 5.9: Den3A-Den15 SS score value results ....................................................... 200
Figure 5.10: M1-M6 DS score value results ............................................................... 204
Figure 5.11: F-J DS score value results ...................................................................... 206
Figure 5.12: Den3A-Den15 DS score value results ..................................................... 227
Table 4.17: Expanded LR results for swatches and yarns .......................................... 123
Table 4.18: Percent accuracy results for denims ........................................................ 126
Table 4.19: Compiled LR results for denims ............................................................... 126
Table 4.20: Expanded LR results for denims .............................................................. 128


LIST OF EQUATIONS

\[ X = TPT + E \quad (2.1) \]

\[ d\text{Mahalanobis} = [(xB - xA)TC - 1(xB - xA)]0.5 \quad (2.2) \]

\[ dx_i = x_i - xTC - 1x_i - x0.5 \quad (2.3) \]

\[ SD_j = l = 1ajtjl2vjl12 \quad (2.4) \]

\[ cSD_j = Xaj, 0.9752 \quad (2.5) \]

\[ OD_i = x_i - P \cdot t_iT \quad (2.6) \]

\[ cOD_j = \text{median}OD_j23 + MADOD_j23 \cdot z0.97532 \quad (2.7) \]

\[ djDx = \gamma OD_jcOD_j + 1 - \gamma SD_jcSD_j \text{ for } j = 1, ..., k \quad (2.8) \]

\[ PrBA = PrB \times PrABPr(A) \quad (2.9) \]

\[ PrHE = PrH \times PrEHPPr(E) \quad (2.10) \]

\[ PrHpEPrHdE = PrEHPPrEhd \times Pr(Hp)Pr(Hd) \quad (2.11) \]

\[ \text{Odds} = \text{Probability of } \omega \text{ occurring}/\text{Probability of } \omega \text{ not occurring} \quad (2.12) \]

\[ \% \text{ Accuracy} = \# \text{True Correct SS Decisions} + \# \text{True Correct DS Decisions}/\text{Total } \# \text{Comparisons} \times 100 \quad (4.1) \]
CHAPTER 1: INTRODUCTION

For decades forensic science has been an integral part of criminal investigations around the world. The first laboratories dedicated to forensic science were created in Europe and now nearly 400 laboratories are accredited under the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) in the United States alone(1). Textile fibers are important pieces of trace evidence because of their ubiquity and presence at most crime scenes.

1.1 Textile Fibers as Trace Evidence

Textile fibers are commonly encountered as forensic evidence during criminal investigations. Different varieties of trace evidence include glass, soil, and paint chips, but what these elements have in common is that they are sought out to provide valuable insight as to the relationship between a suspect and a victim or crime scene. This simple ideology is based on Locard’s exchange principle which states:

> Whenever two objects come into contact there is always a transfer of material. The methods of detection may not be sensitive enough to demonstrate this or the decay rate may be so rapid that all evidence of transfer has vanished after a given time. Nonetheless, the transfer has taken place(2).

This principle allows investigators to adopt the belief that evidence is left behind at a crime scene by a culprit, that it is possible to identify that evidence, and that it may be used to identify the culprit. Many experts believe that textiles offer high evidential value because, although they are produced in large numbers, they are not homogenous, indistinguishable products(3). Even if two garments are purchased from the same store
at the same time, it cannot be assumed that the fibers from each garment are indistinguishable from each other. This is because textile goods are produced in minute batches, the fact that dye batch variation is extremely common, and not all clothing items sold at a particular location can be assumed to be from the same batch(3-6). The implication of these facts is that fiber evidence is capable of being individualized more readily than previously believed given the appropriate tools and protocols.

Forensic fiber comparison is the process by which a questioned fiber is compared to fibers from a known source to determine if an association or match can be made(7, 8). During the 1960s, fibers were proclaimed to be consistent with having a common source and “match” each other if no discrepancies were apparent after color and morphological features were compared beneath a comparison microscope(9). Utilizing polarizing light microscopy (PLM) for fiber examinations did not become common until man-made fibers became more prevalent in modern fashion. Furthermore, it was not until the 1980s that fiber color became the preferred method of comparison over physical morphological features with the introduction of MSP as a forensic technique in England(9). Today, MSP is regarded as the most useful method for nondestructively comparing fiber dyes(10-12).

Common questions asked during fiber examinations include: Is the fiber natural or synthetic? Does the fiber have a generic fiber type? What are the color and/or shade of the fiber? What object did the fiber originate/transfer from? The general scheme for
fiber analysis includes visual examination, microscopical examination, spectroscopic examination, and sometimes chemical analysis when applicable (7, 8). According to the SWGMAT Fiber Subgroup, the forensic analysts must use their own discretion to decide which tests to perform depending on the amount of evidence available and which tests will provide the most discriminatory results (7). Furthermore, it is also up to the analyst to draw conclusions from tests based on their own experience and training and write it up in the final forensic report.

Unfortunately, the process of fiber examination suggested by SWGMAT is susceptible to the subjectivity and bias of the analyst. Furthermore, the human error associated with fiber analysis cannot be quantified (13). When testifying in a criminal trial, the lack of scientific reasoning or evidential support may make it difficult to justify the value or weight of fiber evidence to a judge and jury. This weakness does not lessen the importance of the SWGMAT guidelines, but it does, regrettably, leave room for interpretation of evidence if the results, including terminology and methods, are not explained thoroughly. All scientific methods, including those for forensic analyses, are subject to errors. This does not lessen the value of the conclusions gained from those tests, but it is necessary to understand how likely or how often they will be encountered so that the reliability of the method used can be known (14).
1.2 SWGMAT Guidelines

During the 1980s, with the advancement of nuclear DNA analysis and the realization that false convictions were a reality in the United States, experts in other disciplines of forensic science were faced with adversity: results from forensic examination were no longer blindly accepted as evidence in court proceedings. In order to fortify physical evidence analysis SWGMAT was created. It was first formed under the name of the Technical Working Group for Fibers by Edward Bartick, a former Federal Bureau of Investigation (FBI) employee, in 1994 (15). To this day, the group is made up of bench level scientists working in crime laboratories as opposed to attorneys and academics. The purpose of SWGMAT is to improve trace evidence collection, examination and testimony by creating analysis guidelines, conducting round-robin style research projects, and writing technical papers. The subgroups include those dedicated to fibers, glass, paint, tape, and trace evidence (15, 16).

Members of the Fiber Subgroup have collaborated on developing a set of guidelines for forensic fiber analysis that is made available to the public for free. The original guidelines were published in 1999, but since then revisions to chapters 1-3 were made in 2011 (7, 17-19). Different analytical methods that are described include microscopy of textile fibers, visible spectroscopy of textile fibers, thin-layer chromatography of nonreactive dyes in textile fibers, pyrolysis gas chromatography of textile fibers, infrared (IR) analysis of textile fibers, and fabrics and cordage (17). Here, focus is specifically
placed on the guidelines for visible spectroscopy of textile fibers will be focused on since all other topics fall outside the scope of this work.

The updated examination guidelines clearly state that fiber identification can only be extended to a generic class of fiber, not a specific garment or source. Collection of absorption spectra (240-760 nm) from dyed fibers is considered a nondestructive method for the comparison of fiber dyes. According to the guidelines, averaged spectra or spectral derivatives may be used for spectral comparisons by “overlapping them or by plotting them sequentially on the same graph”(19). When making fiber comparisons, there are two possible outcomes (ignoring an inconclusive result): 1) the questioned fibers are consistent with having come from a known source or 2) they are not consistent with having come from a known source(8). If the ground truth is that the questioned fiber originated from the known source and 1) is reported, or if the ground truth is that the questioned fiber did not originate from the known source and 2) is reported, then no error has occurred. However, it is possible to make false decisions during the course of examination. For example, Type I (an error of false exclusion) and Type II (an error of false association) errors may occur(8, 20, 21). It is possible for Type I errors to contribute to failing to convict a guilty party, whereas Type II errors may contribute to the conviction of an innocent party.

To determine if a positive association is present, the guidelines state that “each questioned fiber spectrum must be compared to the known fiber spectra.”
suggestions on how to make decisions regarding a fiber comparisons are seen in the following excerpt:

A spectral inclusion is when the questioned spectrum falls within the range of the known spectra when considering the curve shape and absorbance values. A spectral exclusion is when the questioned spectrum falls outside the range of the known spectra in either curve shape or absorbance value. An inconclusive result is when there are no significant points of comparison in either the questioned or the known spectra…(19).

Decisions regarding spectral fiber comparisons are based entirely on the shapes of the spectra and are at the mercy of the examiner’s judgment. Outside of calculating the spectral derivative, there is no mention of pretreating the spectral data prior to analysis which is known to aid in the examination of multivariate data such as absorption spectra(22). Additionally, the method described here has some limitations with regard to fibers that have been colored with different dyes of very similar molecular structure that will ultimately exhibit nearly identical absorption spectra(19). The excerpt above discusses the “range of known spectra,” but does not define exactly what the range is. Some analysts may interpret it to mean the range of individual spectra collected from the known source, but perhaps standard deviation curves are more suitable as addressed in section 7.10 of the Fiber Examination Guidelines: “[m]ore scans may be needed if it is necessary to produce a representative mean absorbance curve and standard deviation curves for an individual fiber(19)”. Additionally, it is unclear how an analyst is to handle spectra collected from multiple questioned fibers; this is almost always the case as multiple fibers are collected as evidence from a crime scene. Needless to say, there is much ambiguity regarding the actual fiber comparison guidelines.
Although SWGMAT was created with the intention of improving forensic science practice, some shortcomings were identified and described by Edwards(23). He writes that SWG committees meet irregularly, do not have clear standards for who may gain membership, are not federally regulated, and do not currently attempt to measure how their standards impact the community. Edwards continues to express concern because neither the recommendations nor guidelines agreed upon by SWG committees are enforced in forensic laboratories; it is up to discretion of each laboratory to choose which suggestions to implement if any at all. This is largely problematic because it makes it difficult to monitor the common practices of forensic examination between laboratories. The fact that SWGMAT’s examination guidelines are not mandatory for all forensic analyses is not to say that their work is unimportant or should continue to be overlooked, but the implications are that more collaboration and involvement are needed from the federal government, academic laboratories, and crime laboratories. Although further elaborating on the topic of the problems facing the forensic community are outside the scope of this work, most professional do agree that all crime laboratories should follow the same forensic examination procedures.

1.3 NAS Report

The turning point in forensic science from pure practice to needing to understand the underlying science that forensic examination is built on undoubtedly came with the publication of the NAS report, “Strengthening Forensic Science in the United States: A
Path Forward," in August of 2009(14). The NAS is described in the report as a “private, nonprofit, self-perpetuating society of distinguished scholars” dedicated to furthering science and technology for the greater good of the United States. It has been mandated to advise the government on all things scientific and technical since 1863.

In November of 2005, Congress was directed to authorize the NAS to launch a study on forensic science at the request of both the United States House of Representatives and Senate(24, 25). The Forensic Science Committee, formed by the NAS, was held responsible for examining the current state of all forensic disciplines, assessing their needs and shortcomings, and establishing recommendations for the improvement of forensic science practice. The committee, which constituted members of the forensic science community representing operational crime labs, medical examiners and coroners, and legal experts from across the county, held extensive hearings and deliberated for nearly two years on these matters. In the end, a total of 13 recommendations were endorsed by the committee; they are heavily interconnected and are considered necessary for the improvement of forensic science disciplines(14). Some, but not all, are addressed here.

1.3.1 Recommendations

Throughout these deliberations, and in years prior, a multitude of inadequacies emanating from the forensic science community were brought to light. They include, but are not limited to, forensic methodologies not being founded on scientific principles; lack of regulation of forensic practices among laboratories; unknown error rates of analyses
for various evidence types; and lack of funding, support, and leadership for forensic
disciplines(14, 23, 26-28). Because the forensic science community has been fostered
and maintained by law enforcement agencies, the negative effects of these many
weaknesses have occasionally resulted in false convictions of innocent parties(14, 29,
30). In fact, it is reported that 50% of wrongful conviction cases that were later
overturned by newly available deoxyribonucleic acid (DNA) testing were influenced by
unvalidated or improper forensic science methods and techniques(31). News of this
unfortunate reality bombarded American media during the 1980s and into the 1990s
because advancements in nuclear DNA analysis made it possible to reanalyze genetic
evidence from old criminal cases and exonerate those who were wrongly convicted(14,
26, 30).

Before nuclear DNA technology was utilized for individualization of forensic evidence, it
was extensively researched, and its methods were heavily scrutinized by biologists
within the academic community. This is extremely different than the development of
other forensic disciplines that grew simply out of the need to aid law enforcement
officials in criminal investigations which did not leave room for method validation(14, 28,
32). Nuclear DNA analyses are held as the standard in forensic science because they
exhibit extremely low, non-zero error rates, yet no information regarding the rates of
error of other forensic examinations is known(14). This has fueled much of the
backlash against the forensic science community by law professionals and the public
because all disciplines are expected to deliver the same quality of results. Serious
doubt has been cast on the credibility of the evidence analyzed by non-DNA forensic examinations that often claim discernible uniqueness of evidence because it simply has never been proven(33).

With regard to physical evidence, especially textile fibers, there is one particular recommendation that distributes a heavy load of responsibility onto the forensic science community. Recommendation 3 endorsed by the Forensic Science Committee states, “[r]esearch is needed to address issues of accuracy, reliability, and validity in the forensic science disciplines.” This request can be satisfied by conducting competitively funded peer reviewed research that develops scientifically based forensic methodologies, quantifiable measures of reliability and accuracy of forensic analyses, quantifiable measures of uncertainty, and automated techniques that enhance forensic methods(14). There exists unprecedented pressure on the forensic science community to validate its practices so that they can be held in the same regard as those of the forensic DNA community(8, 23, 26, 33). The research described here aims to work towards fulfilling the needs of the forensic science community, specifically those of forensic fiber analysis, with regard to strengthening the accuracy, reliability, and validity of fiber examination practices.

1.4 Research Goals

The main goal of this research is to advance the state of non-destructive methodology for forensic fiber comparison. It is important to maintain the integrity of the fiber
evidence by employing non-destructive techniques so that the evidence can be revisited at a later time if needed(34). When fiber comparisons fails to discriminate fibers based on physical characteristics, forensic scientists’ next task is to analyze the fiber dyes. For this reason, the research presented here is focused on fiber comparisons using only visible absorption spectral data collected from MSP. Most crime laboratories have access to spectrophotometers that are capable of collecting absorption spectra in the visible range so regardless of treatment of spectral data, the method developed and described here can be universally applied. As previously stated, forensic fiber examination relies heavily on the interpretation of the analyst with regard to physical fiber characteristics and spectral shape; therefore, it is imperative that a completely objective method of analysis is created.

For decades the forensic science community has been faced with the dilemma of determining exactly how to make decisions regarding group member of a questioned fiber: does it belong with the known fibers or not? It was not until the publication of the NAS Report and its recommendations that forensic scientists had a direct path to follow in order to drive the discipline forward. Studies demonstrate that statistical analyses of spectral data collected from dyed textile fibers can greatly enhance forensic fiber examination because those methods offer an objective technique for comparison.

For this work, visible (VIS) absorption spectra collected from the samples of custom dyed fabric swatches, commercially available blue yarns, and denim fabrics. Fiber
comparisons utilizing visual analysis as described by SWGMAT, the score value as a measure of group membership, and the LR as a measure of evidential value are discussed. For the latter of the two methods, performance is measured as a value of % accuracy, or the number of comparisons that were correctly decided given the ground truth.

The purpose of this research is to understand, as best as possible, how well existing method of fiber comparison are able to discriminate between similarly colored textile fibers compared to the methods developed here. Realizing the strengths and weaknesses of new and existing protocols is an important part of helping to fortify the practice of forensic fiber analysis.
CHAPTER 2: BACKGROUND

2.1 Textile Fibers and Their Dyes

Fibers are the smallest components that make up textiles for apparel; household furnishings, upholsteries, and floor coverings; and industrial items such as tents, sails, ropes and cordages(3). Threads are made up of individual fibers and can be used to create woven textiles. Fibers are classified as either natural or man-made, where natural fibers can be vegetable, animal, or mineral fibers and man-made fibers can be made of either synthetic polymers or natural polymers(35-37). As of the year 2000, the most commonly encountered textile fibers are polyester, nylon, acrylic, rayon, and acetate in that order(35). The textile fibers used in this work include cotton (vegetable seed fiber), acetate (natural polymer fibers made from cellulose ester), nylon 6.6 (polyamide synthetic polymer), and acrylic (synthetic polymer with repeating acrylonitrile units)(3).

Fibers are almost always exchanged between a victim and a suspect or a suspect and a crime scene because of their omnipresence; as such, they are often encountered and collected as forensic evidence(38, 39). Depending on the characteristics of the fiber, a textile may be more prone to shed than others. For example, goods made with staple fibers, such as cotton, are more likely to shed than those made with filament fibers, such as silk(3, 38). Obviously, the greater the sheddability, the more likely a fiber is to be found at a crime scene. The degree of sheddability combined with the fact that some
clothing types are more commonly encountered than others contribute to what scientists refer to as fiber frequency(9, 35, 36, 40, 41). Some scientists argue that fibers exhibiting a high fiber frequency are not as valuable as forensic evidence because they are so often collected as evidence(35, 42, 43). However, current literature combats those arguments by showing that some evidential value may be associated with certain types of fiber evidence given the simple facts that fluctuations in manufacturing processes and variations in dye batch formulations occur regularly(4-6, 44). Since many fiber analysts utilize MSP to collect and compare absorption spectra of dyed textile fibers as a part of forensic examination, these slight changes in dye mixtures often manifest themselves in spectral profiles and are capable of aiding in the discrimination of a questioned fiber from known fibers(9, 10, 19, 40, 43, 45, 46).

A dye is a chemical compound used to evoke the visual sensation of a specific color by adhering permanently to a substrate, such as a textile fiber, so that it is able to absorb and reflect complementary wavelengths of the visible spectrum(46, 47). Naturally, there are chemical compounds that are capable of absorbing light only in the ultraviolet (UV) region, but those are not utilized as dyes for textile goods. The wavelengths of light that are able to be absorbed by the dye are based on its molecular structure which is influenced by chromophores and the degree of unsaturation. As the degree of conjugation, or the number of conjugated double bonds, increases the maximum wavelength of absorption also increases(48). Chromophores are the parts of the dye
molecule responsible for absorption of UV-VIS radiation and are often unsaturated moieties\(^{(47, 48)}\).

Fiber dyes fall mainly within nine dye classes: acid, basic, azoic, direct, disperse, reactive, metallized, sulfur, and vat. The chemical composition of the fibers itself and the method of dyeing determine which types of dyes are ideal for application to different fibers. For example, polyamide (nylon) fibers are usually dyed with acid dyes and acrylic fibers are usually dyed with basic dyes\(^{(46)}\). Acid dyes contain sulfonic acid functional groups that interact with the terminal amino groups on the nylon polymer chain allowing ionic bonds to form between the dye molecule and the polyamide. Basic dyes, also known as cationic dyes, are often utilized as ammonium, sulphonium, or oxonium salts and reacted with acetic acid to improve solubility during the dyeing process\(^{(49, 50)}\). Known dyes analyzed in this work include acid, direct, and basic dyes.

As discussed previously, during the production of textiles, slight variations in the dying process can lead to minute differences in the dye composition of the resulting good. Detectable differences in dye batches can be due to alterations in dye mixture, “topping up” of dye batch with additional dye components, over-dyeing of lighter shades to reduce loss, and even sub-contracting to dye houses\(^{(4-6)}\). It has also been reported that the varying nature and amounts of cutting agents, dyeing assistants, and dispersing agents in dye drums can affect dye batch variation\(^{(51)}\). Therefore, although textile goods are mass produced in batches, they are not necessarily indistinguishable from one
another(3). According to Webb-Salter and Wiggins, “[i]f differences can be detected between samples from dye batches, the number of garments that are identical would be limited”(52). The implications of this are that classification of fibers may be taken a step further than merely fiber type and general color. Fibers may not be individualized to a specific garment, but it may be possible to characterize them according to garments or items that originated from a specific batch.

2.1.1 Implications of Chemically Similar Dyes

The human eye perceives light between the wavelengths of approximately 400-700 nm, which corresponds to the visible region of the electromagnetic spectrum. Reactions in the eye-brain system cause humans to experience sensations of light, brightness, and color as a result of the physical stimulus of radiation(47). Although it is extremely sensitive and has been reported that more than six million color shades can be distinguished, color perception of the eye is regarded as very subjective(10, 47, 53, 54). In relation to dyed fibers, metamerism, a phenomenon occurring when two or more colored items produced with different dyes or coloring agents (or different mixtures thereof) appear to be the same color and hue, may occur during forensic analysis(8, 47, 55). The bulk of preliminary fiber examination performed in the crime laboratory requires the visual comparison of the physical characteristics of dyed textile fibers. It is therefore highly likely that metamerism may influence the analyst’s decisions to perform further confirmatory tests.
Utilizing visible absorption spectra collected from dyed textile fibers is known to aid in the discrimination of questioned fibers when other visual tests fail to give conclusive results(8, 36, 45, 56). In a study carried out by Eng et. al., absorbance spectra were collected from metameric blue fibers of cotton, polyester, nylon, silk and wool between 350-800 nm and compared to one another to determine if they could be differentiated. It was concluded that differences in spectral shape for the cotton and polyester fibers were noticeable; these fibers were dyed with different coloring agents but appeared the same color in daylight illumination. The fibers from nylon, silk, and wool materials, on the other hand, were dyed with the same coloring agents but in different ratios. As a result, the spectral shapes were very similar but exhibited differences in relative absorbance intensities at different wavelengths(55).

The shape of an absorption spectrum collected during forensic fiber examination is determined by the absorption of incident radiation by dye molecules. In order for light to be absorbed, the energy of the exciting photon must equal the exact difference in energy between the ground state and one of the excited states of the absorbing species. Molecules have three types of energy states (electronic, vibrational, and rotational) which are different than atoms that only possess electronic states. During visible absorption spectroscopy, the entire visible spectrum is irradiated onto a sample, therefore allowing a multitude of electronic transitions of bonding electrons to occur for different energy differences. As a result, molecular absorptions result in broad, unresolved spectra as opposed to atomic line spectra. The absorption bands are
ultimately an ensemble of closely spaced absorption lines that are not usually resolved unless working with high powered instrumentation (48). Quantum mechanically speaking, an electronic transition requires a large net position overlap of the wave functions in the initial and final states at the instant of transition. In other words, the likelihood of an electronic transition taking place is more likely to happen when there is a greater overlap in vibrational wave functions. The Franck-Condon factor is equal to the square of the vibrational overlap integral and dictates the relative intensities of vibrational bands in absorption spectra (57).

Theoretically, the visible absorption spectra collected from two chemically different dye molecules will exhibit noticeably different shapes. Naturally, the spectral data may be used (probably in conjunction with other characteristics) to easily discriminate fibers from one another during forensic examination. As previously described, MSP in the visible range can even help differentiate between metameric fibers. There are, however, instances where visible MSP may fail to discern between similar dyes or dye mixtures. This occurs when the molecules possess similar molecular structures, degrees of unsaturation, and common chromophores. If the molecules possess very different substituents that are not included in the conjugated system, there probably will not be a large, noticeable difference in spectral shape. Furthermore, the maximum absorption wavelength may not shift to longer wavelengths if the substituents do not increase unsaturation. This is due to the fact that the degree of conjugation remains unchanged and the molecules exhibiting nearly identical Franck-Condon factors. In
Figure 2.1, the dye molecules A and B are chemically identical except for the amide group. The corresponding visible absorption spectra collected from solution are normalized from 0 to 1 and can be seen below. This example shows just how similar resulting spectra are that are taken from a pair of dyes that are very chemically similar.

Figure 2.1: Dye molecules A and B; spectral overlay of visible absorption spectra collected from solutions of A (shown in red) and B (shown in blue)
If and when questioned fibers are analyzed by MSP in a crime laboratory, there is a chance that a misleading conclusion may be drawn from the results of the spectral comparison. The fiber analyst may decide that a questioned fiber is unable to be discriminated from fibers of a known source based on the comparison of the absorption profiles due to the reasons described above. In the event that a discrimination cannot be made, it is suggested by fiber expert Ken Wiggins that further testing be done on the fiber dyes usually by thin layer chromatography (TLC)(49). More involved analytical methods that offer more discriminatory information regarding fiber color include high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), surface enhanced resonance Raman scattering spectroscopy (SERS), and liquid chromatography-mass spectrometry (LC-MS), however, it should be noted that most crime laboratories are not equipped with the instrumentation to perform these kinds of analyses(17, 58-62).

2.1.2 Forensic Fiber Questions

Recent advancements in nuclear DNA analysis have shed light on the serious problem of wrongful convictions in the United States. While there are many factors that may lead to a wrongful conviction, the most influential factor behind faulty eye witness testimony is improper scientific testimony(31). The “CSI effect,” a phenomenon aptly named because of the romanticized crime laboratory culture on television, has further crippled forensic testimony because jurors have come to expect irrefutable forensic evidence in every criminal case brought to court(14, 32). Some, but definitely not all,
forensic analysts have given undue weight to physical evidence as a response to the extreme pressure law enforcement agencies and jurors ultimately place on them(23). It could also be argued that non-existent accreditation requirements and standard practice protocols have led to the demise and current state of forensic science(14). Wrongful convictions, witness impeachment, and crime laboratory failures and shut downs have all cast an enormous shadow of doubt on practically all forensic testimony that is offered in a court of law(14, 29, 63). The simple fact is that the value, reliability, and validity of most methods used during forensic examination is unknown.

Developments made within the forensic science community, including those for fiber analysis, will serve to better assist in criminal investigations by providing more trustworthy conclusions. This will ultimately rebuild faith in forensic testimony, reduce the rate of wrongful convictions, and help ensure that the true criminals are brought to justice. The National Research Council (NRC) strongly recommends that research be conducted to investigate issues of accuracy, reliability, and validity of the numerous forensic examination methods(14). Specifically, scientists need to be able to directly measure these characteristics to include in final case reports in the event expert witness testimony is needed in the future. Fortunately, with regard to method validation, many fiber examination methods are already based on sound chemical principles (MSP and TLC), but scientists still cannot agree on how exactly to draw conclusions from fiber analysis.
Forensic fiber comparison is the comparison of fiber traits and characteristics between an unknown fiber collected from a crime scene (also known as the questioned fiber) and fibers from a known source. The fiber analyst’s main goal is to investigate whether or not the questioned fiber shares a common source with the known fibers. This is usually followed by an assessment of how likely it would be for a completely random fiber to also share a common source with the knowns(38, 40, 64, 65). In other words, if a questioned red polyester fiber was determined to share a common source with the known fibers, the analyst should also determine the possibility of other random, unrelated red polyester fibers sharing a common source with the knowns when compared forensically. In order to determine if the questioned and known fibers share a common source, the analyst must decide which traits or characteristics to compare. Fiber type, width, color, and concentration of delusterants (applicable to man-made fibers only) are all measurements that can be compared; method details will be discussed later. Not all analysts may decide to compare the same type or number of fiber characteristics yet they are all working with the same objective in mind.

Next, the analyst must consider if one known fiber is sufficient to represent the known source or if multiple fibers should be measured to establish a representative sample set(38). As discussed previously, it is incorrect and dangerous to assume that a single fiber can represent an entire textile because factors such as dye batch variation and differences in dye uptake are just two of the factors that can affect fiber color alone. When fiber examination is expanded to include measurements from multiple known
fibers, the analyst must decide which value against which to compare to the questioned fiber. If dealing with univariate measurements, it is suggested that the questioned fiber data be compared to the standard deviation or mean of known fiber measurements(38). Although the implementation of statistical measures is greatly desired in forensic analysis, there are more appropriate and complex methods than those using the standard deviation and mean that will be discussed later. For decades, fiber analysts have relied on personal experience and discretion to decide whether or not fibers share a common source merely by visual comparison of fiber traits(7, 38, 40, 64, 66). Even when analytical methods are used to generate spectral or other chemical data, the actual comparison of this data is done by visual inspection; there is a desperate need to develop more rigorous tests for forensic fiber comparison.

If it is decided that the questioned fiber shares a common source with the known fibers, the analyst’s next objective is to evaluate the chances that a different random fiber may give the same results when compared to the known evidence. In other words, the value of fiber evidence must be determined. Many expert opinions exist which describe how the value of evidence can be evaluated. Some say that evidential value is based on the reliability of the method implemented, how often a procedure draws incorrect conclusions, or even how much influence the outcome of the test will have on the overall case compared to the rest of the evidence being presented. Many scientists support the idea that information concerning fiber frequency in the relevant population is required in order to determine evidential value; that is, are the traits or characteristics
being utilized for fiber comparison rarely encountered or commonly seen(8, 9, 38)? Fiber frequency data has been sought after and collected from various areas for many years, but many factors inhibit large databases from being universally applicable to casework. Target fiber studies, which investigate the likelihood of finding a pair of matching fibers completely by chance, focus only on one type of fiber at a time. Fiber population studies usually examine specific areas or locations and the fibers encountered there; this approach surveys a wider scope but is still limited (67, 68). Fiber frequency information can be difficult to collect, however, a comprehensive textile fiber database would undoubtedly aid in future forensic examination.

Once a decision has been made as to whether or not the questioned fiber shares a common source with the known fibers, the analyst must have a way to evaluate the reliability of the method used. Test performance and method reliability can be evaluated by calculating the number of false inclusions and false exclusions that are reported by a particular method; these measures can also be thought of as error rates when converted to percentages. Finally, the analyst must determine the value, meaning, and implications of the results of a forensic fiber comparison(33, 40). The difficulties in determining evidential value through fiber frequency data have been discussed, yet there is no other specific, direct way to establish value for any type of forensic evidence (other than nuclear DNA) that is currently implemented in a crime laboratory(43). Currently, fiber examiners simply assign evidential value based on experience, but many experts within the forensic community strongly suggest that a
Bayes approach, which is a probability based theorem, be incorporated into forensic analysis, including fiber examination, so that evidential value can be determined with more confidence (9, 69). Using a Bayes approach seems promising to many forensic experts because only sample population data is required to calculate the value of evidence in the form of the LR. These topics will be discussed in greater detail in section 2.3.3.

It is essential for scientists to be able to interpret evidence correctly and establish their credibility as witnesses. Attorneys need to understand how the analyst arrives at the conclusions that are being testified to in court so as to avoid the possibility of a wrongful conviction. Furthermore, it is imperative that jury members interpret the evidence appropriately without giving undue weight to forensic testimony so as to arrive at a well-informed verdict for the defendant.

2.2 Fiber Analysis

The forensic fiber examination performed in crime laboratories is done for the main purpose of eliciting information regarding a crime from the fiber evidence. This information can later be used as a part of expert witness testimony in a court of law or merely as evidence during pre-trial proceedings. For situations involving casework, resources, manpower, and time can be limited. In a survey of 103 laboratories conducted in 2001, it was determined that a general crime laboratory in the United States only employs between 1-3 fiber analysts (70). In 1997, a different survey
distributed to laboratories in North America and Europe showed that a majority of forensic analysts in North American spent less than 25% of their time performing fiber examinations(67). Furthermore, it is well known that most laboratories suffer tremendous backlogs sometimes up to hundreds of thousands of cases(14, 71). In crime laboratories, the fiber analyst does not have the freedom to conduct research to determine which methods are most efficient or reveal the most information from the evidence because of case demands and lack of funding. The analyst must be able to work quickly and employ the protocols enforced by that laboratory.

Forensic fiber research, on the other hand, allows scientists to explore different ideas regarding fiber examination in order to answer those questions that discussed above and so often asked. Many academic research institutions, including the National Center for Forensic Science (NCFS) at the University of Central Florida (UCF), and other federal laboratories are able to apply for and receive competitive funding to conduct such research in order to further develop the current state of forensic fiber examination and to improve the practices employed by the forensic science community. The different types of fiber analysis will be discussed here with the specific intention of discerning between standard forensic fiber practices conducted routinely in crime laboratories and those methods that are pushing the limits of chemistry but are not generally accepted (or even applicable) in most laboratories.
2.2.1 Standard Crime Laboratory Practices

After fiber evidence has been collected from a crime scene, forensic comparison of fiber traces is performed with the intention of fiber identification by fiber type and other characteristics that will be discussed further (8, 72). Fiber identification, determining the generic class of a fiber, is not to be confused with fiber individualization which is reporting that the fiber originated from a particular source (7, 35). It is generally agreed upon by forensic experts that morphological fiber characteristics cannot be used to individualize fiber evidence to its original source, however, there is hope that some analytical technique may have the potential to do this in the future (14). The fiber analyst’s objectives for examining fiber evidence are to distinguish whether it is natural or man-made, assign it to a generic fiber type, assess and determine its color and shade, identify the type of material it could have come from, and, if at all possible, determine where the fiber or textile material was manufactured (8).

According to fiber experts Grieve and Wiggins, the best fiber analysis methods are nondestructive, applicable to very small samples, able to provide maximum amount of discriminatory information, and rapid (40). Although these are excellent points, there is no single fiber examination method that is employed by crime laboratories that can provide satisfactory results while meeting all of these requirements. In fact, it is suggested that multiple tests be performed at the discretion of the analyst to elucidate the needed information to conduct a comprehensive fiber examination. Again, there is no required fiber examination protocol for crime laboratories in the United States, but
the SWGMAT Fiber Examination Guidelines are available so analysts may reference them if they so choose(7, 14, 17-19, 23, 28). These methods of analysis described by SWGMAT and other relevant and important members of the forensic fiber community that are available to fiber examiners are described here.

Microscopical examination of the questioned fiber, particularly with stereomicroscopy and bright-field microscopy, allows the analyst to first make the distinction between natural (plant and animal) and man-made or manufactured fibers(72). Animal fibers or hairs have several characteristics that distinguish them from natural fibers such as the presence of the cuticle (outermost layer of the hair composed of scale-like cells), pigment granules and cortical fusi (air pockets) in the cortex, and the medulla which is the innermost part of the hair(73). Man-made fibers usually possess unique cross-sectional shapes and may contain delustering agents in order to influence the way in which light is reflected off of the material(72). Over half of the fibers used in textiles today are man-made(7, 35). When a man-made fiber has been identified, the analyst uses PLM to observe the various optical properties of the fiber that are not usually prevalent in natural fibers such as isotropic refractive index (RI), RI along the length of the fiber and the width of the fiber, birefringence, sign of elongation, interference colors, and pleochroism(8, 72). IR microspectroscopy and pyrolysis gas chromatography (PyGC) can be used in certain cases to gain information regarding the chemical makeup of the synthetic polymers used to form the fiber itself, however, PyGC is used sparingly because it is a destructive technique(8, 17, 74). IR spectroscopy can only be
used for synthetic fiber identification and not fiber dye identification (36, 46). To gather fiber dye information the fiber analyst will employ methods such as MSP and TLC where MSP is a rapid, non-destructive test and TLC is considered destructive because a dye extraction is required for analysis (8, 36, 46, 47, 49). All of the tests mentioned here are described in even greater detail in the Forensic Fiber Examination Guidelines published by SWGMAT in 1999 and those that were later revised in 2011 (7, 17-19). The type of and number of tests performed on fiber evidence to elicit sufficient information in order to make a fiber identification is completely up to the analyst’s discretion (7, 40). The following chart can be found in Chapter 1 of the SWGMAT guidelines as a tool for analysts to use (7).

Table 2.1: Guide for analysis of fibrous materials

<table>
<thead>
<tr>
<th>Physical Characterization</th>
<th>Optical Characteristics</th>
<th>Chemical Analysis</th>
<th>Color/Dye Analysis</th>
<th>Instrumental Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereomicroscopy</td>
<td>PLM</td>
<td>Solubility</td>
<td>Comparison microscopy</td>
<td>FTIR</td>
</tr>
<tr>
<td>Light microscopy/</td>
<td>Light microscopy/</td>
<td>Staining (natural fibers)</td>
<td>MSP or TLC</td>
<td>SEM-EDS/ XRF</td>
</tr>
<tr>
<td>comparison microscopy</td>
<td>comparison microscopy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEM</td>
<td>Fluorescence</td>
<td>CE</td>
<td></td>
<td>PyGC/ PyGCMS</td>
</tr>
<tr>
<td>Melting point</td>
<td></td>
<td>Raman</td>
<td></td>
<td>XRD</td>
</tr>
<tr>
<td>Physical test (dry twist,</td>
<td></td>
<td></td>
<td></td>
<td>Raman</td>
</tr>
<tr>
<td>ashing, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Once the identity of the questioned and known fibers have been established, comparisons between them are performed (40). Fiber comparison of physical characteristics takes into consideration the fibers’ color; diameter; presence, amount, size, shape, and distribution of delusterants; cross-sectional shape; any surface characteristics such as inclusions, damage, and adhering debris; and indications that
the fiber was textured, processed, or printed(8). Comparison of optical and fluorescent
properties such as the presence or absence of fluorescence, excitation wavelength
producing the maximum wavelength, and the color and intensity of fluorescence under
various excitation conditions are integral for forensic analysis when applicable. There
exists no rule for the analyst to know when all tests have been exhausted or when to
end fiber analysis. As stated before, the analyst is responsible for choosing the best
and number of analytical tests for fiber evidence. When recording results from
casework in the final forensic report, the value of the fiber evidence is not always
reported, but when it is, it is reported based on the experience and opinion of the
analyst(38, 40). The question now is: how can the true value of fiber evidence be
quantified and objectively reported?

2.2.2 Forensic Fiber Dye Research

Color analysis of textile fibers is arguable the most important aspect of forensic fiber
examination. It is likely the most influential features when determining if a questioned
fiber shares a common source with known fibers. Although there are numerous
morphological and optical properties that are considered as points of comparison during
examination, these characteristics are mostly only present in man-made fibers and not
natural fibers. Therefore, qualitatively and quantitatively measuring the color of dyes
and the ratios of the dyes used to impart color onto textile materials is a major aspect of
forensic fiber research conducted in laboratories throughout the world.
Although not an instrumental technique, TLC is a chemical technique used to separate the individual parts of extracted fiber dyes by allowing a solvent system to separate the dye components on a stationary phase. This technique has the ability to differentiate between optically isomeric dye pairs that appear the same color by visual inspection and often exhibit similar visual absorption profiles, yet have different chemical structures(8). Based on the degree of chemical affinity of the dye components for either the solvent system or the stationary phase, the dyes will migrate along a porous silica gel and the analyst is able to classify them into specific dye classes(17, 49). MSP and TLC are considered complimentary techniques to one another where MSP is a non-destructive technique used to collect spectral information directly from a dyed fiber sample and TLC is able to provide dye class information from a dye extraction(17). Even though these two methods are rapid and routinely used in crime laboratories, they are limited in the sense that very darkly colored fibers are not easily analyzed via MSP and suitable dye extracts cannot be made from too lightly colored fibers for TLC(8, 46, 47). Neither technique is able to provide definitive information for fiber analysts to identify a specific dye or dyes within a mixture.

It is evident from literature review that research utilizing HPLC, and variations of it, is the most explored technique for forensic fiber dye analysis behind UV-VIS MSP. In a study conducted by Huang et al., the authors specifically utilize HPLC in conjunction with a UV-VIS absorption detector in series with electrospray ionization mass spectrometry (ESI-MS) to analyze dye extracts from textile fibers(62). If dye extracts are collected
from known and questioned fibers that are dyed with chemically similar dyes it is very likely that the chromatograms from standalone HPLC with a UV-VIS absorption detector will exhibit indistinguishable retention times due to coeluting species and similar absorption profiles. The mass spectral information provided by MS is essential in order to confidently identify the dye or dyes used to color the fibers based on the formation of molecular ion fragments.

In a similar project, Petrick et al. employed a similar HPLC UV-VIS spectroscopy ESI-MS setup to develop a method for analysis of basic and disperse dye extracts collected from “casework-sized samples”(75). Until the advent of ESI, fiber dyes could not readily be analyzed by MS following separation by LC because scientists could not easily create gas phase ions of the dye extracts and because high flow rates were not suitable for the high vacuum environment required by the mass spectrometer. Eventually, researchers determined that when a “soft” or “mild” ionization technique such as ESI was used instead of an electron beam, molecular fragmentation of dyes can be easily and more readily induced prior to the ions entering the mass spectrometer(75, 76).

Since dye extractions are mandatory for chromatographic and subsequent mass spectral analysis of textile fibers, it is important for fiber analysts to determine if an extraction is feasible for the type and amount of sample present. Some dye classes are not suited for easy extraction due to the way in which the dye molecules have bonded to the fiber substrate, and in some cases the number of fibers collected from a crime
scene is too small to perform an adequate dye extraction(8). This issue was explored by Tuinman et al. when research was conducted to develop a method to analyze fiber dye extracts collected from submillimeter nylon fibers by direct infusion into ESI-MS(77). Real-world nylon fiber samples were analyzed, and it was determined that winding (thread) extracts were able to be differentiated from one another without first being separated by chromatography. This study also explores collision-induced dissociation (CID) experiments in order to elucidate more significant structural information about the precursor ions. The authors decided, however, that even if the chemical structure cannot be determined from the CID spectrum (which was the case in some instances), it may still serve as a “fingerprint” that is unique to the sample and can later be used for identification purposes.

In a study conducted by Huang et al., HPLC ESI-MS was utilized to examine dye pairs that were known to exhibit nearly identical UV-VIS absorption spectra due to their chemical structures being highly similar(61). Seven dye pairs were analyzed: Acid Green 25 and 27, Acid Red 4 and 8, Acid Red 14 and 73, Basic Red 9 and Basic Violet 14, Disperse Blue 3 and 14, Disperse Red 1 and 13, and Solvent Red 26 and 27 and Sudan III. (Disperse Blue 3 and 14 are also investigated in the original research presented here.) After it was demonstrated that the individual dyes within each pair could be differentiated by comparison of the mass spectra, the method was tested on 10 commercially available “red” cotton items that appeared indistinguishable by microscopic examination. The mass spectra of these 10 items showed that none
shared a common source. Methods utilizing HPLC ESI-MS have been proven to be suitable for the forensic analysis and discrimination of textile fiber dyes.

While there are many scientists working to improve upon the various disciplines of forensic science by performing validation studies by incorporating analytical and chemically established methods, a small fraction of those projects utilize chemometric techniques. Chemometrics will be explained in much greater detail later, however, for now it is sufficient to know that the application of statistics to multivariate chemical data may be done as a part of forensic analysis. Although the amount of research that employs chemometric methods has greatly increased since the publication of the NAS Report just over five years ago, forensic scientists still have much to test, optimize, verify, and quantitate.

As a part of the original dissertation research carried out by Rex in 2009, parallel factor analysis (PARAFAC) and multivariate curve resolution alternation least squares (MCR-ALS) was utilized in order to compare highly similar excitation emission matrices (EEM) formed from room-temperature fluorescence spectra of dyed textile fibers(78). Both chemometric techniques were evaluated on their effectiveness as more robust discrimination techniques for forensic fiber comparisons. Among the samples analyzed, fiber extracts from a pair of nylon fibers dyed with Acid yellow 17 and 23 respectively were able to be discriminated from one another when MCR-ALS was employed even though the fibers were visually indistinguishable.
White also published original thesis research out of UCF applied where various statistical and chemometric techniques were applied to the absorption spectra and mass spectra collected from a number of different dyed textile mediums(79). Visible absorption spectra were collected from red acrylic yarn fibers, red cotton fibers, blue acrylic yarn fibers, and dyed fabric samples, while mass spectra were collected from dye extracts from the blue acrylic yarn fibers and the dyed fabric samples. Among the many statistical and chemometric techniques employed, the ones that are most relevant to the current research being presented are PCA and Discriminant Analysis (DA). Distinct groups were visualized after PCA was performed on fibers from the five blue acrylic yarns was performed: Yarn H, Yarns F and J, and Yarns G and I represent the groups. It was hypothesized that the reason for Yarns F and J and Yarns G and I clustering together was because each pair shared a common manufacturer and the yarns were dyed different shades of the same color (blue). MS was later able to discriminate them from one another despite each pair clustering together as a result of PCA. Furthermore, visible spectra collected from fibers of the four dyed fabric samples (Disperse Blue 3 and 14 and Basic Green 1 and 4) were also subjected to PCA, and it was reported that the clusters for each sample are clearly separated from one another. Although samples from each dye pair clustered separately, discrimination cannot be confirmed based solely on visual recognition of natural groupings. Dyed textile fiber analysis using PCA, which will be discussed in much greater detail later, has also been
implemented in graduate work by Liszewski, Szudlarek, Reichard, and Appalaneni (80-83).

2.2.2.1 Chemometrics Applied to Spectra from Colored Traces

Chemometric analysis of chemical data generated from forensic examination of evidence is able to offer the structure and objectivity needed for comparing forensic traces and determining whether or not they share common sources. There is a substantial amount of literature dedicated to the application of multivariate statistics to the analysis of spectral data collected from colored traces including inks, paints, and dyed textile fibers. In 2003, Thanasoulias et al. performed multivariate chemometric analysis on the visible absorption spectra collected from 50 blue ballpoint pen inks from five different brand names in order to determine if forensic discriminations between them were possible (54). PCA was used as a method for outlier removal after feature reduction was performed by cluster analysis through the K-means method. Finally, the new variables formed by PCA were subjected to DA and it was determined that when pen inks from different brands were compared to one another, 100% discrimination was achieved.

The next year at the European Fibres Group Annual Meeting in Prague, Czechoslovakia, a paper was presented on the “Forensic Discrimination of Dyed Textile Fibers using UV-VIS and Fluorescence Microspectrophotometry” (34). Morgan et al. analyzed spectra from both dyed and undyed cotton, polyester, acrylic, and nylon fibers
using PCA and Linear Discriminant Analysis (LDA) in order to test the method’s
discrimination ability. Leave-one-out cross validation was performed for the fiber
comparisons. The cross-validated classification based on the PCA projections for
cotton, acrylic, nylon, and polyester samples were 70%, 80%, 67.5%, and 90%
respectively, while for the LDA projections it was 100% for each fiber group.

In a study conducted by Adam et al., the forensic classification and individualization of
25 different black ballpoint pen inks using PCA applied to the corresponding UV-VIS
absorption spectra(84). When PCA was performed on the full data set including two
dye standards, the results show five distinct clusters: A, B, R, C, and D. Groups A, B,
and R which were clearly separated were removed from the dataset, and PCA was
repeated on the remaining data in order to elucidate more defined clusters. The authors
use the term “discriminated” to describe clusters, or groupings of data points from the
pen ink samples, that are visually separated from one another, yet there is no mention
of a distance metric that was used to objectively determine the true discrimination of the
sample; this type of testing does not further advance the forensic science community if
subjective visual examinations are the deciding factor for discrimination of evidence.

A few years later in 2011, researchers from Indiana University-Purdue University
Indianapolis (IUPUI) utilized multivariate statistics for the forensic discrimination of dyed
hair color(85). Over 50 red hair dyes were applied to standard hair bundles from one
individual, and the spectra were collected via MSP. The chemometrics techniques
employed for this study include Agglomerative Hierarchical Clustering (AHC), PCA, and DA. Overall, Barrett et al. found that although the techniques described were not capable of individualizing the dyed and undyed hair samples, but the results strongly support future research that is focused on validating forensic analyses of dyed hair using chemometric techniques. Knowing the limitations of any forensic technique provides valuable information to the rest of the community.

Most recently, Appalaneni et al. was able to achieve single fiber identification through cluster analysis of EEM data(86). Some of the samples studied included two dye pairs (Acid Blue 25 and 41 and Direct Blue 1 and 53) that were used to dye swatches of spun nylon 351 and cotton 400, respectively. Training data was collected from a single fiber from each of the four dyed swatches. For comparisons, ellipses boundaries were calculated from 3x the standard deviation of the training cluster and spectral data collected from individual fibers, threads, and regions of the corresponding dyed swatches were projected into the training data space. If the projection fell inside the ellipse, then it was classified with that training set, but if it fell outside of the ellipse then further testing was done using an $F$-test to confirm its identity. The results showed that data collected from individual fibers, threads, and regions dyed with AB25 were all classified correctly using the $F$-test with 99% confidence and none were misclassified as AB41. The same was true for the reverse situation except for one thread that failed to classify as neither AB41 nor AB25. For the second dye pair, the results were more involved: 80% of fibers from DB1 were correctly classified with two fibers that failed to
classify as neither DB1 nor DB53, and all threads and regions dyed with DB1 were correctly identified at 99% confidence. As for DB53, 90% of individual fibers, 100% of threads, and 80% of regions were classified as such while none were incorrectly classified as DB1.

Chemometric analysis is accepted as a powerful analytical tool that has great potential when applied appropriately within various forensic science disciplines. The literature and research discussed here, although specific to the analysis of colored traces, demonstrates the popularity and demand for multivariate statistical methods for the comparison of physical evidence. In the future, researchers aim to develop methods utilizing statistical techniques that are simple to implement, require straightforward explanations in court, and answer those questions previously described: how reliable is this method? And how valuable is this evidence?

2.3 Chemometric Methods

Current spectroscopic, chromatographic, and mass spectral methods produce data of high dimensional, multivariate data. It has been reported that subtle differences in large data sets are nearly impossible to identify visually and that real differences between samples of a multivariate data set may not be manifested simply as the presence or absence of a single peak in a spectrum or chromatogram(22, 87). Chemometrics is known as a chemical discipline that utilizes statistical methods for two main purposes: to design optimal procedures and experiments and to extract the most relevant and
important chemical information by analyzing chemical data(88-90). Chemometric analysis of multivariate chemical data is currently being heavily explored in various science disciplines and forensic chemistry is no exception(22, 46, 85, 88, 91). Much of forensic research has already begun to utilize various statistical methods to better make forensic comparisons and to test the limits of what physical evidence can reveal to criminal investigators. Although a long way from being used within a crime laboratory, these methods seek to explore the questions and concerns set for by the NAS committee regarding validity of current forensic methods and the true value of forensic evidence.

Specifically when performing MSP on fiber evidence, it is very possible that the decision of whether or not a spectral inclusion or exclusion can be determined may depend on subtle fluctuations in the spectral data, and it is possible for features to be masked by noise(22). This, combined with the issues of metamerism and the fact that chemical dye molecules can exhibit similar Franck-Condon factors, desperately begs for a more robust, objective method of comparison for absorption spectra collected from similarly dyed textile fiber evidence. The method developed should not be dependent on analyst bias gained from experience or the circumstances surrounding the case at hand. Chemometrics is believed to be the avenue to develop such a method; however, very few chemists (let alone forensic scientists) have a substantial background in multivariate statistics. The chemometric methods used in this work are described below.
2.3.1 Principal Components Analysis

One of the most widely used chemometric techniques for computing latent variables from a large, multivariate data set, and one of the methods employed in this work, is PCA(88, 91, 92). It involves an abstract mathematical transformation of the original data matrix where the rows are samples (individual spectra or chromatograms) and the columns are variables (wavelengths or time)(91, 92). PCA is characterized as an unsupervised learning technique, which means that the method does not require any prior knowledge of class membership of the data set(88, 92). Unsupervised learning techniques can help visualize patterns in the data (groupings or clusters) to better understand the behavior of the samples and determine what other statistical techniques should be applied. Perhaps most important feature of PCA is that it is able to reduce the dimensionality of the data in order to make it more manageable for subsequent analyses(22, 85, 91). In other words, PCA can be used to extract the most important information from a data set by projecting it into lower dimensional subspaces where a new coordinate system is created, and the new axes describe the maximum amount of variability in the data(22, 85, 90). Figure 2.2 shows data points (black dots) in two dimensions where the red dotted line describes the maximum variance of those points, and the green dotted lines describe the variance in the orthogonal direction.
The new lines or axes are called principal components (PC) and are considered latent variables. A unique property of PCs is that they are all orthogonal to one another (88). According to Massart et al., PCs are “uncorrelated linear functions of the original variables” (89). The value of a PC is referred to as a score (88). The orientation of these PCs is determined such that they retain the maximum variation along them and minimize the variation around them (22, 88, 89).

PCA describes an original data set, \(X\), as:

\[
X = TP^T + E
\]  

(2.1)
where $X$ has the dimensions $n$ rows (samples) by $m$ columns (variables), $T$ equals the scores matrix with the dimensions $n$ rows by $a$ columns ($a$ is equal to the number of principal components), $P^T$ equals the loadings matrix with the dimensions $a$ rows by $m$ columns, and $E$ equals the error in measurements, from instrumental noise, and other sources not explained by the product $TP^T$ (91, 92). Figure 2.3 demonstrates this relationship.

![Image](image_url)

**Figure 2.3: Approximate reconstruction of $X$-matrix. Note: Adapted with permission from Introduction to Multivariate Statistical Analysis in Chemometrics (p. 62) by K. Varmuza and P. Filzmoser, Boca Raton, FL: Taylor and Francis Group, LLC. Copyright 2009(88)**

In Equation 2.3.1.1, the product $TP$ is known as a model or an approximation of the original data set, given by $X_{appr}$, without the error $E$. The scores in $T$ are considered linear combinations of the loadings, coefficients of the original variables, and the original variables themselves (88, 89). In other words, the loadings describe the influence or
weight of the original variables on the scores. Each row in the loadings matrix, $P^T$, contains a vector of loadings pertaining to its corresponding PC and describes a direction in the variables space(88). According to Varmuza, “[a]ny pair of latent variables defines a projection of the $m$-dimensional variable space on to a plane given by the loading vectors and the scores describe the projection coordinates”(88). Loadings are also known as the coordinates of the corresponding eigenvector(22, 89). An eigenvalue describes the variance along the corresponding eigenvector(89). The first PC is always defined by the largest eigenvalue where all subsequent PCs are described by lesser and lesser eigenvalues.

2.3.2 Distance Metrics

Distance metrics are used to assess the distance and similarity between from two samples; a large distance equates little similarity between objects and vice versa. The greater the similarity between objects the more likely it is that they originate from the same group. One of the distance metrics used in this work is the Mahalanobis distance, which takes into account the distribution of data points within the variable space, specifically the PCA space, and is considered independent from the scaling of the variables. Traditionally, the Mahalanobis distance is used for outlier detection of multivariate data and is characterized by the covariance matrix; it is particularly useful when different variances and correlations exist between variables(88, 93). The Mahalanobis distance is calculated from the following equation:
where two objects are defined by vectors $x_A$ and $x_B$ with components/variables $x_{A_1}, x_{A_2}, \ldots, x_{A_m}$ and $x_{B_1}, x_{B_2}, \ldots, x_{B_m}$ and $C$ is the sample covariance matrix. The Mahalanobis distance can also be calculated from each observation to the center of the data as:

$$d(x_i) = [(x_i - \bar{x})^T C^{-1} (x_i - \bar{x})]^{0.5} \quad (2.3)$$

where $x_i$ is an object vector, $i$ is the number of variables in the data matrix, and $\bar{x}$ is the arithmetic mean vector.(88)

When the Mahalanobis distance is used as a measure between the center of PC space and a data point, $x$, it can also be defined as the score distance:

$$SD_j = \left[ \sum_{l=1}^{a_j} \frac{t_{jl}^2}{v_{jl}} \right]^{1/2} \quad (2.4)$$

where $a$ is the number of PCs forming the PC space, $t_{jl}$ are the components of the scores $t_j = (t_{jl}, \ldots, t_{ja_j})^T$ and $v_{jl}$ are the largest eigenvalues for $l = 1, \ldots, a_j$ in the $j^{th}$ group(88). When determining group membership of an object, the decision is made based on a cutoff threshold. In this case, the threshold, or cutoff value, for the score distance is based on the chi-squared distribution with $a$ degrees of freedom at the 97.5% quantile:

$$c_{SD_j} = \sqrt{\chi^2_{a_j,0.975}} \quad (2.5)$$
To understand the utility of this metric, imagine a 3-dimensional space where the known scores only require a two component solution; that is, only the first two PCs are required to represent a majority of the total variance of the original data set. Therefore, the known scores lie within (or close to) a plane in 3-dimensional space. Figure 2.4 shows the known scores as black circles and their positions with respect to the plane shown in red:

![Figure 2.4: Known scores on a 2-dimensional plane in 3-dimensional space](image)

The small blue dot represents the center of the position of the plane, but not necessarily of the 3-dimensional space. From the image it can be seen that some of the scores lie slightly above and below the plane, but generally are associated with the plane itself. When a questioned sample is projected into this space, the score distance
(Mahalanobis distance) can be used to describe its position relative to the rest of the known data. Figure 2.5 shows the questioned sample as a solid black circle sitting above the plane:

![Diagram showing Mahalanobis distance](image)

**Figure 2.5: Questioned score projected onto 2-dimensional plane**

When the questioned sample is projected onto the surface of the plane of the 2-dimensional PC space at a 90° angle (green line), the distance from the projection point of the questioned score to the blue dot is the score distance (black line). If the score distance is less than the cutoff value then the questioned score may be considered a member of the group of known scores, but if it is greater than the cutoff then it is considered an outlier, or a non-member.
Looking back at Figure 2.5 it is evident that even though the score distance may be small for the questioned sample, its distance above the plane is quite large in relation to the score distance. Consider a situation where the question sample was projected into the same PC space where its score distance was rather large as shown in Figure 2.6:

![Figure 2.6: Questioned score projected onto 2-dimensional plane](image)

In this case, the score distance is much larger than the distance between the point and the plane. For this reason, it would be incorrect to only consider the score distance when deciding the membership of a questioned sample, therefore, statisticians also use the orthogonal distance. This distance is given by the length of the orthogonal projection of the score onto the plane (green line) and is calculated from:

\[ OD_i = \| x_i - P \cdot t_i^T \| \]  (2.6)
where $x_i$ is the $i^{th}$ object of the centered data matrix, $P$ is the loadings matrix using $a$ PCs, and $t_i^T$ is the transposed score vector of object $i$ for $a$ PCs(88). The double brackets are indicative of the Euclidean norm being taken for each sample in the original data matrix(89). The cutoff value for the orthogonal distance is given by:

$$c_{OD_j} = \left( \text{median} \left( OD_{ij}^2 \right) + \text{MAD} \left( OD_{ij}^2 \right) \cdot z_{0.975} \right)^{\frac{3}{2}} \quad (2.7)$$

where $MAD$ is the median absolute deviation and $z_{0.975}$ is the 97.5% quantile of the standard normal distribution (a value of 1.96)(88).

Traditionally these two distance metrics are used in a technique called Soft Independent Modeling of Class Analogy (SIMCA) that is used to classify a sample among a much larger group of samples. It is possible for the sample to classify to a single group, multiple groups, or none of the groups present and this method is referred to as a soft modeling technique because more than one outcome is possible(88). For this research, however, this method is used in order to determine if a questioned fiber sample can or cannot be classified to a single group: the known fiber samples. The classification is dependent on the calculation of the score value given by:

$$d^P(x) = \gamma \left( \frac{OD_{ij}}{c_{ODj}} \right) + (1 - \gamma) \left( \frac{SD_{ij}}{c_{SDj}} \right) \quad \text{for } j = 1, \ldots, k \quad (2.8)$$

where $\gamma$ is an optimization parameter on the interval [0,1] that gives weight to both the score and the orthogonal distances for classification. A Coomans plot (Figure 2.7) can be utilized to visualize the placement of a questioned object with respect the cutoff.
values for each of the distances in order to determine whether or not it can be classified with the knowns(88).

Figure 2.7: Coomans plot showing cutoff values for both score and orthogonal distances. Note: Adapted with permission from Introduction to Multivariate Statistical Analysis in Chemometrics (p. 212) by K. Varmuza and P. Filzmoser, Boca Raton, FL: Taylor and Francis Group, LLC. Copyright 2009(88)

2.3.3 Likelihood Ratio

Determining and reporting the evidential value of forensic comparisons is a difficult task for forensic experts. Since the recommendations from the 2009 NAS report were published, many scientists have explored the use of Bayesian theory to calculate evidential value because statisticians believe it is able to serve as a model for interpreting forensic evidence(94). Bayes’ theorem, which can be seen in Equation 2.9,
is read: the probability of event B occurring given A is equal to the probability of event B multiplied by the probability of event A occurring given B divided by the probability of A(21).

\[ Pr(B|A) = \frac{Pr(B) \times Pr(A|B)}{Pr(A)} \quad (2.9) \]

If hypothesis, H, is substituted for B and evidence, E, is substituted for A, then the equation becomes:

\[ Pr(H|E) = \frac{Pr(H) \times Pr(E|H)}{Pr(E)} \quad (2.10) \]

In the forensic context, there are two competing hypotheses: prosecution hypothesis \((H_p)\) and defense hypothesis \((H_d)\). If Equation 2.10 is rearranged in order to accommodate both hypotheses, it becomes:

\[ \frac{Pr(H_p|E)}{Pr(H_d|E)} = \frac{Pr(E|H_p)}{Pr(E|H_d)} \times \frac{Pr(H_p)}{Pr(H_d)} \quad (2.11) \]

For forensic comparisons, the prosecution hypothesis supports the belief that the questioned evidence shares a common source with the known evidence, while the defense hypothesis supports the belief that the questioned and known evidence do not share a common source. Support of \(H_p\) usually implies that the defendant is guilty because it assumes that the traces recovered from the crime scene (questioned evidence) share a common source with the defendant's possessions (known evidence). Following that logic, support of \(H_p\) would then imply that the defendant is not guilty. The first term is known as the posterior odds, the middle term is known as the LR, and the last term is known as the prior odds. The prior and posterior odds deal with evaluating
the probability of guilt or innocence based on the evidence presented, whereas the LR evaluates the probability of encountering evidence given the innocence or guilt of the defendant. The forensic scientist should never share an opinion on the prior or posterior odds since they should only be concerned with the evidence at hand.

Probability is a standard for uncertainty of an event occurring, or in this case a hypothesis, being true or not; it is written as \( \Pr(H) \)\(^{(95)}\). The odds of an event occurring are defined as\(^{(21)}\):

\[
\text{Odds} = \frac{\text{Probability of } \omega \text{ occurring}}{\text{Probability of } \omega \text{ not occurring}} \tag{2.12}
\]

where \( \omega \) is the event. The prior odds are in favor of the defendant’s guilt without being conditioned on the evidence, \( E \); that is, the odds of guilt are determined without consideration of knowledge of the evidence. The posterior odds of guilt are what the jury decides after being presented all the evidence. The LR determined for forensic evidence is the probability of encountering evidence given that the defendant is truly guilty divided by the probability of encountering evidence given that the defendant is truly innocent. LRs may be calculated through the use of evidence population data (sometimes through the use of databases where available) that describe the similarity of characteristics, variation among and between sample, frequency of observed features, and any correlation between the features of the multivariate data (i.e. absorption spectra)\(^{(53)}\). This information is often difficult to collect in its entirety, therefore, forensic
scientists are forced to make necessary assumptions about what is acceptable population data.

Probability values may range on the interval $[0,1]$, but LR values may range on the interval $[0,\infty]$ (95). According to Evett, et al., support of an individual forensic comparison may be assigned, in varying degrees, to either of the two competing hypotheses using the verbal scale shown in Table 2.2 (96):

**Table 2.2: Evett’s verbal scale for likelihood ratios**

<table>
<thead>
<tr>
<th>LR</th>
<th>Verbal Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LR \leq 0.0001$</td>
<td>Very Strong Support ($H_d$)</td>
</tr>
<tr>
<td>$0.0001 &lt; LR \leq 0.001$</td>
<td>Strong Support ($H_d$)</td>
</tr>
<tr>
<td>$0.001 &lt; LR \leq 10$</td>
<td>Moderately Strong Support ($H_d$)</td>
</tr>
<tr>
<td>$0.01 &lt; LR \leq 0.1$</td>
<td>Moderate Support ($H_d$)</td>
</tr>
<tr>
<td>$0.1 &lt; LR \leq 1$</td>
<td>Limited Support ($H_d$)</td>
</tr>
<tr>
<td>$1 &lt; LR \leq 10$</td>
<td>Limited Support ($H_p$)</td>
</tr>
<tr>
<td>$10 &lt; LR \leq 100$</td>
<td>Moderate Support ($H_p$)</td>
</tr>
<tr>
<td>$100 &lt; LR \leq 1000$</td>
<td>Moderately Strong Support ($H_p$)</td>
</tr>
<tr>
<td>$1000 &lt; LR \leq 10000$</td>
<td>Strong Support ($H_p$)</td>
</tr>
<tr>
<td>$10000 &lt; LR$</td>
<td>Very Strong Support ($H_p$)</td>
</tr>
</tbody>
</table>

The purpose of reporting the LR in a forensic report or as expert testimony is not to say conclusively if questioned evidence comes from the same course or can be discriminated from the known evidence, its purpose is to aid in the interpretation of the value of that forensic comparison. The LR helps the expert relay the weight of the forensic comparison to the judge and jury.
CHAPTER 3: EXPERIMENTAL

3.1 Instrumental Parameters

For these experiments, visible absorption spectra were collected from textile fibers and later subjected to chemometric analyses. The microscope used here is an Olympus System Microscopes model BX51 with an Olympus U-SPT V clamp attached to the observation tube. This model is considered a polarizing microscope although it was not utilized in this capacity. The objective used for all fiber measurements was a 40X Olympus UPlanFL P-series objective with a numerical aperture of 0.75, a working distance of 0.51 mm, an infinity-corrected tube length, and a correction for standard 0.17 mm thick coverslips. The ocular contributes a magnification of 10X by itself so the total magnification of the system is 400X. A CRAIC Technologies QDI 302 microscope spectrophotometer was coupled to the microscope to collect the visible absorption spectra from the individual textile fibers. The spectrometer contains a thermoelectric-cooled Sony ILX511 CCD array detector, and the grating element contains 600 lines/mm blazed at 500 nm. A 1.3 megapixel digital imaging system equipped with a FireWire CCD color camera (model DFK 41AF02) was employed to visualize the exact area on the sample that was measured at any given time through the computer monitor. Visible absorption spectra were collected within the wavelength range of 400-725 nm using CRAIC Technologies CRAIC MSP Data Acquisition software with a spectral resolution of approximately 0.66 nm. The collection parameters were set to average 50
scans per acquisition. Integration time was determined by the auto-optimize feature within the software.

3.2 Samples Sets

3.2.1 Dyed Fabric Swatch Samples

The first of three sample sets studied in this work included three pairs of custom dyed fabric swatches individually dyed with chemically similar dyes. The dye pairs were chosen for this study because they exhibited similar absorption spectra and molecular structure. Dyestuffs were obtained from Sigma-Aldrich (St. Louis, MO, US) and Acros (Geel, Belgium) and sent to Test Fabrics, Inc. (West Pittston, PA, USA) for custom dyeing. Table 3.1 displays the dye details:

<table>
<thead>
<tr>
<th>Swatch Label</th>
<th>Dye Name</th>
<th>Fabric Type/Style</th>
<th>$\lambda_{max}$ (in nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Acid Blue 25</td>
<td>Spun Nylon 6.6; Style 361</td>
<td>600 (in water)</td>
</tr>
<tr>
<td>M2</td>
<td>Acid Blue 41</td>
<td></td>
<td>599 (in water)</td>
</tr>
<tr>
<td>M3</td>
<td>Disperse Blue 3</td>
<td>Acetate Satin; Style 105B</td>
<td>640, 594 (in 50% ethanol)</td>
</tr>
<tr>
<td>M4</td>
<td>Disperse Blue 14</td>
<td></td>
<td>640, 594 (in 50% ethanol)</td>
</tr>
<tr>
<td>M5</td>
<td>Basic Green 1</td>
<td>Spun Acrylic; Style 864</td>
<td>625 (in 50% ethanol)</td>
</tr>
<tr>
<td>M6</td>
<td>Basic Green 4</td>
<td></td>
<td>614 (in water)</td>
</tr>
</tbody>
</table>

The molecular structures of the dye pairs can be seen below in Figure 3.1:
Figure 3.1: Molecular structures of Acid Blue 25 and 41 (top pair), Disperse Blue 3 and 14 (middle pair), and Basic Green 1 and 4 (bottom pair)
All swatches weighed 10 g each and were dyed with 0.3 g of dyestuff according to in-house dyeing procedures established at Test Fabrics, Inc. This set of dyed fabric swatches were created in order to represent a simplified version of a fiber analyst's most difficult task: discriminating between two fibers that originated from sources dyed with chemically similar dye pairs. It is uncommon for textile goods to be dyed with only a single dye; however, method development was contingent on the creation and testing of this sample set.

3.2.2 Blue Yarn Samples

The second sample set that better represented case samples was made up of five skeins of commercially available blue yarn. Yarn samples were purchased from craft retailers in the greater Orlando, FL, and were chosen because they were considered to be visually indistinguishable. Table 3.2 shows the yarn information:

<table>
<thead>
<tr>
<th>Yarn Label</th>
<th>Brand and Style</th>
<th>Fiber Type</th>
<th>Color/Shade</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Bernat Satin</td>
<td>100% Acrylic</td>
<td>Admiral 04110</td>
</tr>
<tr>
<td>G</td>
<td>Caron Simply Soft Quick</td>
<td>100% Acrylic</td>
<td>Navy 0005</td>
</tr>
<tr>
<td>H</td>
<td>Red Heart Super Saver</td>
<td>100% Acrylic</td>
<td>0387 Soft Navy</td>
</tr>
<tr>
<td>I</td>
<td>Caron Simply Soft</td>
<td>100% Acrylic</td>
<td>DK Country Blue 9711</td>
</tr>
<tr>
<td>J</td>
<td>Bernat Satin Sport</td>
<td>100% Acrylic</td>
<td>Marina 03110</td>
</tr>
</tbody>
</table>

Note that yarns F and J and yarns G and I are marketed and sold by the same brand, yet are listed as being different shades of blue. Information pertaining to the identity of the dyes used to color the yarns is unknown.
3.2.3 Denim Fabric Samples

The third and final sample set contained seven denim fabrics (not denim pants) purchased from fabric retailers in the greater Orlando, FL, area. Table 3.3 lists the information for these samples:

Table 3.3: Denim fabrics information

<table>
<thead>
<tr>
<th>Denim Label</th>
<th>Description</th>
<th>Fabric Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Den3A</td>
<td>Dress Denim by Oakhurst textiles</td>
<td>100% Cotton</td>
</tr>
<tr>
<td>Den4</td>
<td>Denim Basic SBL WSH DNM</td>
<td>100% Cotton</td>
</tr>
<tr>
<td>Den5</td>
<td>Bottomweight Crosshatch denim</td>
<td>100% Cotton</td>
</tr>
<tr>
<td>Den6</td>
<td>Basic Denim, Indigo Wash DNM</td>
<td>100% Cotton</td>
</tr>
<tr>
<td>Den8</td>
<td>Fashion Denim, 7.5 oz D BL CRSHT DNM SPRING BOT</td>
<td>100% Cotton</td>
</tr>
<tr>
<td>Den9</td>
<td>Denim-Basic Indigo WSH DNM 10 oz</td>
<td>100% Cotton</td>
</tr>
<tr>
<td>Den15</td>
<td>Cotton Bttmwt solid, L BL ICE WASH</td>
<td>100% Cotton</td>
</tr>
</tbody>
</table>

Again, information pertaining to the identity of the dyes used to color the denim fabrics is unknown. Denim cottons are notoriously difficult to discriminate based on the forensic comparison of absorption spectra. For this reason, they were studied at great length for this research.

3.3 Sampling Methods

3.3.1 Dyed Fabric Swatches

For the dyed fabric swatches, individual fibers were collected from threads from three different areas on the swatch. Figure 3.2 demonstrates where the fibers were sampled from:
Areas 1, 3, and 5 correspond to the upper left corner, center, and bottom right corner of the swatch, respectively. Swatches were sent to NCFS precut so no knowledge of the warp or weft directions exists. Swatches were, however, received with fabric type information written at the bottom of each swatch in black marker; this was used as a reference to determine horizontal and vertical directions of the threads. A total of 10 fibers were sampled from each swatch. For M1, M2, M5 and M6, fibers in both directions were non-delustered (no delustering agents present), while vertical fibers of M3 and M4 were non-delustered and horizontal fibers of both were delustered. Because of this, 10 fibers were sampled from each direction specifically for M3 and M4.
Fibers ranged from 1-2 cm in length and were mounted between a glass microscope slide and coverslip in immersion oil. Fibers from each swatch were analyzed one at a time. A total of 15 absorption spectra were collected along the length of each of the 10 fibers (from each swatch). A dark scan and reference scan were collected before each absorption spectrum was collected. The spectral measurements were labeled with a code based on the swatch it was sampled from (M1-M6), the area it was sampled from (A1-A5), the direction of the thread (D1: vertical, D2: horizontal), the thread it was collected from (T1, T2, etc.), the fiber number within the thread (F1, F2, etc.), and the scan number (S1, S2, etc.). Once all spectra were collected using the CRAIC MSP Data Acquisition Software, the files were exported in .txt file format for subsequent data compiling prior to chemometric analysis. Each absorption spectrum contained 920 wavelength variables.

Each set of spectra collected from a single fiber was compiled into a file where the first column contained the wavelength variables and each subsequent column contained the absorption values from each spectral scan. This was accomplished by a compiler software written in Matlab (The MathWorks, Inc., Natick, MA, USA). All compiled files were saved in .CSV format.

3.3.2 Blue Yarns

The blue yarns dataset is unique because fibers from all five skeins were sampled and analyzed previously by White in 2009(79). In a sense, the visible absorption spectra
collected by White were recycled in this research such that they were subjected to
different statistical tests. The sampling methods described for the blue yarns are
slightly different than those described previously for the dyed fabric swatches. Fibers
were sampled from each of the five yarns F-J, and each fiber was cut into three
segments. The microscope used to collect the visible absorption spectra was a Nikon
Eclipse E600 POL with a Nikon C-CU universal system condenser. All measurements
were collected with a 40X objective from the Nikon Plan Fluor Series that had a flat field
and fluorite aberration corrections, a numerical aperture of 0.75, a working distance of
0.72 mm, an infinity-corrected tube length, and a correction for standard coverslips
(0.17 mm-thick). The spectrometer used to collect the spectral measurements was an
Ocean Optics USB-4000-UV-VIS miniature fiber optic spectrometer equipped with a
Toshiba TCD1304AP linear CCD array detector. The spectrometer had a grading of
600 lines/mm blazed at 300 nm (79). After fibers were mounted between a microscope
slide and coverslip, 10 measurements were collected from each fiber segment using
SpectraSuite software (Ocean Optics, Inc.) in the range of 400-700 nm and a spectral
resolution of approximately 0.2 nm. Integration time for the collection of absorption
spectra was determined using the auto-integration function within the software.
According to White, “[s]moothing was accomplished by averaging several scans per
spectrum, and the number averaged varied by experiment” (79).

Because the current work is not concerned with making forensic comparisons between
different parts of the same fiber, the spectra collected from all segments of the same
fiber were treated collectively. A total of 30 absorption spectra were considered representative of a single fiber. The code used to label the spectra is based on the yarn it was sampled from (F-J), the fiber it was collected from (1-5), and the spectral scan number (1, 2, etc.). Each absorption spectrum contained 1533 wavelength variables. Data files used for subsequent chemometric analyses were saved in .txt format and were compiled manually so that the first row contained the wavelength variables and all subsequent rows contained the absorption values from each spectral scan. Each row was identified by a number (1-5) indicating which fiber the spectrum was collected from.

3.3.3 Denim Fabrics

The denim fabrics analyzed were cut from large rolls of material at the fabric retailers; the resulting pieces of denim measured approximately 50"x5". Figure 3.3 shows a generic diagram of how the denim fabrics were sampled:

Figure 3.3: Generic sampling areas on denim fabrics (1, 2, 3, and 4)
The seven denim fabrics studied here were constructed with both undyed (white) cotton threads and dyed cotton threads. An example of fabric construction can be seen below in Figure 3.4:

![Figure 3.4: Close up image of Den3A surface](image)

Fibers from only the dyed threads were analyzed. Two fibers were sampled from each area so that a total of eight fibers were sampled from each denim fabric. Fibers were mounted in immersion oil between glass microscope slides and coverslips. Following the parameters described in section 3.1, a total of 15 visible absorption spectra were collected along the length of each of the fibers. Spectral measurements were labeled with a code based on the fabric they were sampled from (Den3A-Den15), the area they were collected from (A1-A4), the fiber that was analyzed (F1, F2), and the scan number (S1, S2, etc.). All spectra were collected with the CRAIC MSP Data Acquisition
Software, and files were exported in .txt format for subsequent data compiling before chemometric analysis. Each absorption spectrum contained 920 wavelength variables.

Each set of spectra collected from a single fiber was compiled into a file where the first column contained the wavelength variables and each subsequent column contained the absorption values from each spectral scan. This was accomplished by a compiler software written in Matlab (The MathWorks, Inc., Natick, MA, USA). All compiled files were saved in .CSV format.

3.4 Visual Analysis of Spectral Variance

The representative average spectra were calculated from fibers collected from each material and plotted sequentially on the same graph to visualize the variance in absorption and to identify any irregularities in spectral profiles.

3.5 Visual Comparisons of Absorption Spectra

Visual absorption spectra collected from dyed textile fibers were compared visually according to the SWGMAT protocol described in the examination guidelines described in section 1.2. Code was written in R to perform spectral comparisons(97). Spectral comparisons were made by overlaying the representative average spectrum from the questioned fiber onto the standard deviation curves calculated based on the spectra collected from the known fibers. Within sample comparisons, referred to here as same sample (SS) comparisons, were performed to determine the rate of false exclusions
(Type I error) based solely on visual examination. Between sample comparisons, referred to here as different sample (DS) comparisons, were performed to determine the rate of false inclusions (Type II error) based solely on visual examination.

Since none of the samples originated from casework, it was decided to assign certain fibers to roles of questioned and known fibers. For SS comparisons, a single fiber from a particular sample (i.e. M1, G, or Den8) was assigned as the questioned fiber while the remaining fibers were assigned as the known fibers. The questioned fiber’s representative average spectrum was then compared against the upper and lower limit standard deviation curves calculated from the known fibers’ spectra. After this comparison was made, the next fiber within that same sample dataset was assigned as the questioned fiber while the remaining fibers were assigned as the known fibers. The comparisons continued until every single fiber from the sample was assigned as the questioned fiber. After all SS comparisons for a particular sample were made, a new sample was selected. Comparisons are performed until all SS comparisons were made for each sample in all sample sets.

For DS comparisons one pair of samples is investigated at a time. The fibers belonging to the first sample in the pair (i.e. M1) were designated as the known fibers. The fibers belonging to the second sample in the pair (i.e. M2) were designated as the questioned fibers. Comparisons were performed between the representative spectrum from each questioned fiber and the upper and lower limit standard deviation curves calculated from
all of the known fibers’ spectra. The DS comparisons continued until all questioned fibers were compared individually to the set of known fibers in the pairing. Before selecting another sample pair, the first sample in the pair was reassigned as the questioned and the second sample in the pair was reassigned as the known; all DS comparisons were performed. Then all other sample pairs were selected and compared. Sample pairs for the dyed fabric swatches were chosen based on dye pair information. DS comparisons for blue yarns were made for yarns that shared a common manufacturer. All combinations of denim fabric pairs were compared.

3.6 Fiber Comparisons Using Score Value

The score and orthogonal distances were used to calculate the score value in order for it to be used as a measure of group membership for forensic fiber comparisons. Comparisons were considered as either SS or DS comparisons as previously described. R code was written to first calculate the representative average spectrum for each fiber in the input dataset. Various data pretreatment options are written into the code for normalization, nm, and include summing all values in each spectrum to 1, making the maximum value in each spectrum equal to 1, normalize each spectrum to unit length vector, scale all values from 0 to 1, and zero each spectrum to minimum value. In the code, these options are given by the variables “s”, “m”, “v”, “01”, and “0”, respectively. The only option and default method for data centering is mean centering (given by center=”mc”).
For SS comparisons, the input dataset, given by variable x, includes spectra collected from fibers of a particular sample (i.e. M3D1, G, or Den15). For DS comparisons, the input dataset included two sets of spectra, given by variables x and y, collected from a pair of samples. Next, PCA was executed on the known spectra using only the \textit{svd} function in R. An option for outlier removal based on either the Mahalanobis distance or the robust Mahalanobis distance is available, but none was performed for these calculations. Once PCA was performed on the known spectra, the number of PCs that were represented 95% of the variance were retained for the remainder of the calculations.

The questioned fiber sample was then projected into the PC space, and the score value was calculated for the questioned sample according to Equation 2.8. The Mahalanobis distance cutoff value is calculated based on the chosen quantile value, which is set to 0.975 and can be interpreted as the 97.5% quantile. The optimization parameter, $\gamma$, was tested at several values for each sample set (i.e. dyed fabric swatches, blue yarns, or denim fabrics). Fiber comparisons were conducted using the same logic as the spectral comparisons where SS and DS comparisons were performed to evaluate Type I and Type II errors respectively. These calculations were done using the \textit{chemometrics, calibrate}, and \textit{lattice} packages in R.
3.7 Fiber Comparisons Using Likelihood Ratios

Likelihood ratios were calculated from modelled probability distribution functions based on measures of similarity (referred to as scores) from pairwise comparisons of representative average spectra. The input files for these calculations were created differently than the ones for calculations using score value. Here, an input file contains representative average spectra from all fibers that the user wishes to compare. The file may contain spectra from a pair of materials or spectra from several materials. For the experiments described here, dyed swatches and yarns were treated together while the denim fabrics were analyzed separately. It was decided to treat the denims separately because of the knowledge that other traditional methods of fiber analysis often fail to discriminate between different samples. By analyzing the denim spectra separately, method optimization was performed without the influence of other samples.

Pairwise comparisons of absorption spectra were quantified using measures of similarity. Prior to conducting the comparisons, the representative absorption spectra are pretreated according to the options described in section 3.6. The only difference is that there is an addition center method; “a”, which stands for autoscaling (where the variance is set equal to 1) is available. The resulting scores from comparisons among representative average spectra from each sample and from comparisons between spectra from different samples (not to be confused with scores from PCA) were calculated after data pretreatment. The code is written to randomly assign 80% of the scores to a training set which is used to establish the normal SS and DS probability
density function models. Figure 3.5 shows an example of the normal probability density functions.

Figure 3.5: Example of normal probability density functions

The blue curve represents the normal density function modelled after all SS comparisons while the red curve represents the normal density function modelled after all DS comparisons. The remaining 20% of SS and DS correlations were used as data in the test set in order to calculate the LRs. Figure 3.6 demonstrates how a LR is calculated based on a test score value and the corresponding SS and DS probabilities.
Figure 3.6: Example of LR calculation from normal probability density functions

The ratio of the SS probability (value at blue x) to the DS probability (value at red x) at a specified test score (yellow x) was calculated as the likelihood ratio. These were calculated for every value within the test data set. The Evett Scale was used to assign verbal statements of the likelihood ratios(96). The verbal scale is utilized to assign a degree of support to either the defense hypothesis or the prosecution hypothesis to any given comparison within the test set. These calculations were done using the pROC and caret packages in R.
CHAPTER 4: RESULTS AND DISCUSSION

4.1 Visual Inspection of Spectral Variance

The natural variance in spectral profiles of the three sample sets was investigated by visual inspection. The average representative spectrum from each individual fiber was used to create spectral overlays for the different fabrics in each sample set. The spectra seen here are not normalized in any way; the averages were calculated from raw data.

4.1.1 Spectral Variance in Dyed Fabric Swatches

In Figure 4.1, the average representative spectra from fibers of each swatch are overlayed onto one another.
Figure 4.1: Spectral overlays for each dyed fabric swatch, M1-M6; fiber1=red, fiber2=blue, fiber3=green, fiber4=purple, fiber5=pink, fiber6=light blue, fiber7=orange, fiber8=lilac, fiber9=gray, and fiber10=black
Generally, the spectral shapes of those representative spectra from M1 are similar. Between the wavelength range of approximately 575-665 nm, the spectra exhibit the most variation in relative absorption, but this may be attributed to differences in dye uptake and fiber twisting as previously described. The spectral overlay for M2 exhibits similar variations within the same wavelength range. When looking at the orange spectrum from M1, it can be seen that it deviates from the general shape between 400-475 nm. Also, it is obvious that the light blue spectrum from M2 has an overall lower absorbance when compared to the rest of the spectra; since absorbance is dependent on concentration according to Beer’s law, this event is probably due to the fact that this particular fiber was not dyed as deeply as the others.

The spectral overlays for M3D1 and M4D1 exhibit less vertical spread than that exhibited by M3D2 and M4D2. As explained in section 3.3.1, M3 and M4 were sampled differently due to the presence of delustrants only in one direction of the fibers (D2). M3D1 and M4D1 fibers are expected to show less variability in the spectral profiles since there are no delustrants (light-scattering particles) present in the fiber matrix. The spectral overlays for M3D2 and M4D2 show much more variation in absorption between spectra since delustrants are incorporated into the acetate polymer.

The profiles for M5 and M6 exhibit the least amount of absorption variability when compared to all the other dyed swatches. This pair of fabrics was analyzed within a month of being dyed, whereas the other dyed fabrics were analyzed several years after
being dyed. It is possible that exposure to sunlight or other environmental conditions caused a certain degree of photobleaching for pairs M1-M2 and M3-M4. These changes to the fibers themselves may have attributed to the spectral inconsistencies described before. Lastly, the black spectrum from M5 has the lowest overall absorbance when compared to the other spectra; this is most likely due to a lower concentration of dye being present in this particular fiber.

4.1.2 Spectral Variance in Blue Yarns

The spectral overlay plots for yarns F, G, H, I, and J can be seen in Figure 4.2
Figure 4.2: Spectral overlays for each blue yarn, F-J; fiber1=red, fiber2=blue, fiber3=green, fiber4=purple, fiber5=pink
The spectra profiles of both yarn pairs F-J and G-I are mostly similar in shape. F and J exhibit three different maxima, while G and I both show two maxima. The average spectra from H also exhibits two maxima, however, the peaks are broader than those present in the plots for G and I. Just by visual comparison it can be seen that all five plots show approximately the same degree of absorbance variability. It is possible that the differences in maximum absorbance of each spectrum were due to variations in dye uptake along the fiber itself. Because of manufacturing processes, the fibers used to make the threads for a skein of yarn might be dyed in multiple batches at different times and/or locations which could possibly cause noticeable spectral variations in the absorption profiles collected from individual fibers.

The yarns used in these analyses were purchased directly from craft stores, therefore, no information was obtained concerning how long or to what extent the samples were exposed to elements such as direct sunlight or other environmental conditions. If they were, then it is possible for dyes to fade or become altered; any changes that did occur would manifest themselves in the shapes of the spectral profiles.

4.1.3 Spectral Variance in Denim Fabrics

The representative average spectra from each of the denim fabrics can be seen in Figure 4.3.
Figure 4.3: Spectral overlays for each denim fabrics, Den3A-Den15; fiber1=red, fiber2=blue, fiber3=green, fiber4=purple, fiber5=pink, fiber6=light blue

Generally, the shapes of the spectra from each of the denim fabrics are similar in shape to one another. Just based on visual inspection, it appears that the absorption bands at
\( \lambda_{\text{max}} = 660 \text{ nm} \) for Den4 and Den15 are narrower than the absorption bands for Den3A and Den5 at the same wavelength. The lilac spectrum in Den3A, the green spectrum in Den5, and the blue spectrum in Den8 are the only three instances where there is a noticeable deviation from the main group of spectra due to large differences in absorbance intensity. The rest of the spectra from the remaining denim fabrics exhibit approximately the same amount of vertical spread in the spectral profiles.

These denim fabrics exhibit very similar spectral shapes because they are most likely dyed with different combinations of the same or similar dyes. Although there are a multitude of different denim brands and washes, denims usually exhibit the classic medium blue to dark indigo color. Blue cotton denims have always posed a problem for forensic fiber analysts employing methods that compare the visible absorption spectra, and these spectral plots show just how similar the profiles are.

This study of spectral variation among individual samples of the dyed fabric swatches, blue yarns, and denims demonstrates that some fibers exhibit more variability than others. Furthermore, fibers containing delustrants exhibit the greatest variation among spectra, and fibers from yarns that were dyed more recently showed less variability in spectral profiles than those dyed several years prior to testing. When performing visual comparisons of absorption spectra, the analyst must be aware of the natural variance already present among the fibers. It is likely that if known fibers exhibited a lot of inherent variability in the corresponding absorption spectra that a comparison of a
questioned fiber from a completely different source would be falsely included with the
knowns; the known spectra variability might mask the fact that the questioned fiber does
not actually share a common source with the knowns.

4.2 Visual Comparisons of Absorption Spectra

Spectral comparisons were made to emulate the types of comparisons suggested by
the SWGMAT Fiber Subgroup as described in sections 1.2 and 3.5. Note, the standard
deviation curves created from the known spectra are calculated based on the mean of
the known spectra +/- 2x the standard deviation (accepted as the 95% confidence
interval) since it is unclear from the SWGMAT Fiber Examination Guidelines how
exactly to establish these boundaries. All of the SS and DS visual comparison result
plots can be seen in Appendix A.

4.2.1 SS Comparisons

Within sample comparisons modelled after the proposed SWGMAT protocol were
performed for swatches M1-M6, yarns F-J, and denims Den3A-Den15. The ground
truth in these cases was that the questioned and known fibers chosen for the
comparisons were sampled from the same swatch, yarn, or denim, and they are not
expected to discriminate (result in a spectral exclusion) from one another. If any
spectral exclusions did occur, they were considered false exclusions that contribute to
Type I error. The term “exclusions” was used here because the SWGMAT Fiber
Subgroup encourages analysts to use this terminology when writing final forensic reports. All samples are subjected to R code written to plot the spectra and standard deviation curves for SS comparisons. The following excerpt from the updated “Ultraviolet-Visible Spectroscopy of Textile Fibers” chapter of the Fiber Examination Guidelines was used by the researcher to decide if a spectral inclusion or exclusion occurred for each comparison(19):

A spectral inclusion is when the questioned spectrum falls within the range of the known spectra when considering the curve shape and absorbance values. A spectral exclusion is when the questioned spectrum falls outside the range of the known spectra in either curve shape or absorbance value. An inconclusive result is when there are no significant points of comparison in either the questioned or the known spectra…

4.2.1.1 SS Comparisons for Dyed Fabric Swatches

A total of 80 SS comparisons were made for dyed fabric swatches by visual analysis. Table 4.1 shows a summary of the false spectral exclusion results for each of the dyed fabric swatches:

Table 4.1: Summary of false spectral exclusions for M1-M6

<table>
<thead>
<tr>
<th>Sample</th>
<th>M1</th>
<th>M2</th>
<th>M3D1</th>
<th>M3D2</th>
<th>M4D1</th>
<th>M4D2</th>
<th>M5</th>
<th>M6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number SS Comparisons</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Number Spectral Exclusions</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% False Exclusions</td>
<td>6.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The false exclusion rate of 6.25% means that the questioned fiber was falsely excluded from the known fibers 6.25% of the time despite the ground truth (the questioned and
known fibers shared a common source). Because these exclusions are based entirely on visual comparisons made by the researcher, the spectral shapes were investigated thoroughly. In Figure 4.4, all five of the SS comparisons that were falsely excluded can be seen.
Figure 4.4: M1-M6 SS false spectral exclusions
The first two spectral exclusions were both present in M1. For M1A5D2T1F1 v M1 Knowns, the questioned spectrum (shown in red) falls just barely outside of the standard boundary between the wavelength range of approximately 500-560 nm. This particular spectrum also differs in shape at the blue end of the spectrum compared to the standard deviation curves. In the second comparison, M1A5D2T2F4 v M1 Knowns, the questioned spectrum falls outside of the standard deviation boundary over the wavelength range of 580-660 nm. Although the general shapes of both of the questioned spectra are very similar to the standard deviation boundary profiles, it appears that the relative intensities are what caused the questioned spectra to ultimately be excluded from the knowns.

The only spectral exclusion present for M2 is when M2A3D2T3F1 was assigned as the questioned fiber. Looking at this spectral comparison, it can be seen that nearly the entire questioned spectrum lies below the lower limit of the standard deviation curve. As in the previous case, the general spectral profile coincides with that of the boundary profiles, but the intensity of the absorbance values is what caused this spectrum to be discriminated. In the forensic context, false exclusions may contribute to the misdirection of a criminal investigation that does not pursue the truly guilty party. It is possible that a bench scientist would conclude the M2A3D2T3F1 v M2 Knowns comparison to be a spectral inclusion regardless of where it falls in relation to the standard deviation curves, but this decision would be at his or her discretion.
The last two spectral exclusions occur for M3A3D2T3F4 v M3D2 Knowns and M5A5D2T3F1 v M5 Knowns, respectively. Note that for the former comparison, the questioned spectrum differs only in the shortest (400-410 nm) and the longest (660-725 nm) portions of the visible spectrum. This was an extreme case where the majority of the spectrum lies within the standard deviation boundaries but, in an attempt to be conservative, was considered a spectral exclusion. The fifth and final spectral exclusion recorded for the dyed swatches occurs when M5A5D2T3F1 was selected as the questioned fiber. The questioned spectrum deviates slightly from the standard deviation boundary at approximately 600-655 nm.

4.2.1.2 SS Comparisons for Blue Yarns

A total of 25 SS comparisons were made for blue yarns by visual analysis. Table 4.2 shows a summary of the false spectral exclusion results for each of the blue yarns:

Table 4.2: Summary of false spectral exclusions for F-J

<table>
<thead>
<tr>
<th>Sample</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number SS Comparisons</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number Spectral Exclusions</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% False Exclusions</td>
<td>4.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The false exclusion rate means that the SS comparisons were wrongly discriminated 4.00% of the time. The only exclusion that occurs is very unique in that it was discriminated by the profile shape - not because it fell outside of the standard deviation
boundary. Figure 4.5 displays the singular false exclusion when F1 is assigned as the questioned fiber.

![F1 vs Known Spectra](image)

**Figure 4.5: F1 v F Knowns SS false spectral exclusion**

At the blue end of the spectrum (400-410 nm) the questioned spectrum does not exhibit the same shape as compared to the shapes of the standard deviation curves. Figure 4.6 shows that the average spectrum for F1 (red) was actually an anomaly at that end of the spectrum when compared to the average spectra from fibers F2-F5 (black).
Again, it is possible that a fiber expert could still declare this comparison a spectral inclusion based on the rest of the spectral profile; however, when following strict interpretation of the SWGMAT protocol, this comparison would be considered a spectral exclusion.

4.2.1.3 SS Comparisons for Denim Fabrics

A total of 56 SS comparisons were made for denim fabrics by visual analysis. Table 4.3 shows a summary of the false spectral exclusion results for each of the denim fabrics:
Table 4.3: Summary of false spectral exclusions for Den3A-Den15

<table>
<thead>
<tr>
<th>Sample</th>
<th>Den3A</th>
<th>Den4</th>
<th>Den5</th>
<th>Den6</th>
<th>Den8</th>
<th>Den9</th>
<th>Den15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number SS Comparisons</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Number Spectral Exclusions</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>% False Exclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.36</td>
</tr>
</tbody>
</table>

Following strict interpretation of the SWGMAT rules, SS denim fabric comparisons were falsely excluded 5.36% of the time. The three exclusions occur for comparisons Den3AA4F2 v Den3A Knowns, Den5A2F1 v Den5 Knowns, and Den6A4F2 v Den6 Knowns. Figure 4.7 shows the spectral comparisons:
In the first two cases, the red questioned spectrum clearly falls below the lower limit of the standard deviation curve. The questioned spectrum in the third case (Den6A4F2) crosses the boundary at opposite ends of the spectrum only.
4.2.2 DS Comparisons

Between sample comparisons modelled after the SWGMAT protocol were performed for swatch pairs (M1-M2, M3D1-M4D1, M3D2-M4D2, and M5-M6), yarns pairs (F-J and G-I), and all pairwise denim fabric comparisons. The ground truth in these cases was that the questioned and known fibers chosen for the comparisons are sampled from different swatches, yarns, or denims, and they were expected to discriminate (spectral exclusion) from one another. If any spectral inclusions (non-discriminations) did occur, they were considered false inclusions that contribute to Type II error. The term “inclusions” was used here because the SWGMAT Fiber Subgroup encourages analysts to use this terminology when writing final forensic reports. All samples were subjected to R code written to plot the spectra and standard deviation curves for DS comparisons. The following same excerpt from the updated “Ultraviolet-Visible Spectroscopy of Textile Fibers” chapter of the Fiber Examination Guidelines from section 4.2.1 was used to decide if a spectral inclusion or exclusion occurred for each comparison.

4.2.2.1 DS Comparisons for Dyed Fabric Swatch Pairs

A total of 80 DS comparisons for dyed fabric swatch pairs were made by visual analysis. The first sample in the pair is the known and the second sample in the pair is the questioned. Recall that each individual fiber from the questioned sample is what was used to make the comparisons. Table 4.4 shows a summary of the false spectral inclusion results for each of the swatch pairs.
Table 4.4: Summary of false spectral inclusions for swatch pairs

<table>
<thead>
<tr>
<th>Sample Comparisons</th>
<th>M2vM1</th>
<th>M1vM2</th>
<th>M4D1vM3D1</th>
<th>M3D1vM4D1</th>
<th>M4D2vM3D2</th>
<th>M3D2vM4D2</th>
<th>M6vM5</th>
<th>M5vM6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Comparisons</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Number False Inclusions</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The sample comparisons in the above table were labeled such that the first sample represents the known and the second sample represents the questioned. For example, in M2 v M1, M2 was the known sample and M1 was the questioned sample. This table shows that there was a 0% rate of false inclusion when M1-M6 DS comparisons were performed. Some of the comparisons made showed the questioned spectrum falling outside of the standard deviation boundary (Figure 4.8), but other questioned samples were excluded by differences in spectral shape.

Figure 4.8: Examples of spectral exclusion due to boundary violation
For example, comparisons involving M3 and M4 (both directions), the profile of the upper and lower limit boundaries exhibited two distinct absorption maxima at 594 nm and 640 nm. Figure 4.9 displays two DS comparisons between M4A3D2T3F3 v M3D2 Knowns and M3A1D1T1F1 v M4D1 Knowns.

![Figure 4.9: Example of peak inversion for M1-M6 DS comparisons](image)

When M3 is the known (left plot), it can be seen that the 594 nm peak exhibited a higher absorbance than the 640 nm peak. On the other hand, the 594 nm peak exhibits a lower absorbance than the 640 nm peak when M4 serves as the known (right plot). Therefore, the inversion of relative peak heights of the two maxima was used to exclude M3 fibers from M4 fibers and vice versa.
4.2.2.2 DS Comparisons for Blue Yarn Pairs

A total of 20 DS comparisons were made for blue yarn pairs by visual analysis. Table 4.5 shows a summary of the false spectral inclusion results for each of the yarn pairs.

Table 4.5: Summary of false spectral inclusions for yarn pairs

<table>
<thead>
<tr>
<th>Sample Comparisons</th>
<th>JvF</th>
<th>FvJ</th>
<th>IvG</th>
<th>GvI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Comparisons</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number False Inclusions</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number Inconclusive Decisions</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>% False Inclusions</td>
<td></td>
<td></td>
<td>85.0</td>
<td></td>
</tr>
</tbody>
</table>

According to these results, DS yarn comparisons are falsely included 85.0% of the time. This is a significant jump from results of the previous sample set. Figure 4.10 shows just two examples of false inclusions for comparisons G1 v I Knowns and F2 v J Knowns.
Figure 4.10: Two examples of false inclusion results for yarn pairs

The shape and position of the red questioned spectrum in relation to the standard deviation curves in both cases does not allow for spectral exclusion. Nearly all of the DS yarn comparisons failed to exclude from one another. During the visual inspection of spectral variance described in section 4.1.2 it was noted that the yarns exhibited approximately equal amounts of vertical spread in the absorbance intensities. Because of the amount of distance between the upper and lower limits of the standard deviation curves, this makes the possibility of a false inclusion more of a reality. It is possible that if the gap between the boundaries was decreased then the test would be more exclusive. Calculation the boundaries based on 1x the standard deviation instead of 2x may improve the yarn results (lower false inclusion rate), but the false exclusion rate will likely increase as a result. Furthermore, this method needs to be widely applicable for a
multitude of forensic fiber comparisons so optimizing for a single sample set does not benefit other types of comparisons.

There were two interesting cases for yarn pair comparisons where an inconclusive result was decided. These two cases (J2 v F Knowns and J3 v F Knowns) are displayed in Figure 4.11.

![Figure 4.11: Two examples of inconclusive results from yarn pairs](image)

From the plots, a peak is seen at approximately 630 nm. In both cases, the questioned spectrum does not exhibit a prominent peak at the same wavelength even though the general shape of the profile is consistent with that of the known. It is unknown whether the lack of peak definition would be cause for a spectral exclusion despite the spectra
falling within the boundaries. A fiber analyst posed with this challenge while working in a crime laboratory may make an inconclusive decision due to lack of spectral characteristics from which to draw a definitive spectral decision.

4.2.2.3 Pairwise DS Comparisons for Denim Fabrics

A total of 336 pairwise DS comparisons were made for all denim fabric pairs by visual analysis. Table 4.6 shows a summary of the false spectral inclusion results for the denim fabrics.
Table 4.6: Summary of false spectral inclusions for denim fabrics

<table>
<thead>
<tr>
<th>Sample Comparisons</th>
<th>Number Comparisons</th>
<th>Number False Inclusions</th>
<th>Total Number Comparisons</th>
<th>% False Inclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Den4vDen3A</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den5vDen3A</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den6vDen3A</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den8vDen3A</td>
<td>8</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den9vDen3A</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den15vDen3A</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den3AvDen4</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den5vDen4</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den6vDen4</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den8vDen4</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den9vDen4</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den15vDen4</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den3AvDen5</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den4vDen5</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den6vDen5</td>
<td>8</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den8vDen5</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den9vDen5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den15vDen5</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den3AvDen6</td>
<td>8</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den4vDen6</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den5vDen6</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den8vDen6</td>
<td>8</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den9vDen6</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den15vDen6</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den3AvDen8</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den4vDen8</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den5vDen8</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den6vDen8</td>
<td>8</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den9vDen8</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den15vDen8</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den3AvDen9</td>
<td>8</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den4vDen9</td>
<td>8</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den5vDen9</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den6vDen9</td>
<td>8</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den8vDen9</td>
<td>8</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den15vDen9</td>
<td>8</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den3AvDen15</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den4vDen15</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den5vDen15</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den6vDen15</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den8vDen15</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Den9vDen15</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

336  58.0%
The rate of false inclusions for the DS comparisons is 58.0%. Based on the similarity seen in the different absorption profiles of the denim fiber these results were not surprising. In Figure 4.12 two examples of false spectral inclusions can be seen for Den3AA2F1 v Den8 Knowns and Den6A3F2 v Den9 Knowns.

![Figure 4.12: Two examples of false inclusions for DS denim comparisons](image)

As seen in section 4.1.3, the denim spectra did not exhibit many absorption characteristics that offered much to aid in discrimination. Furthermore, there was a significant gap between the standard deviation curves that could have allowed for a questioned spectrum to be incorrectly characterized as a spectral inclusion. It is possible that fewer false inclusions might occur if the curves are calculated from 1x the standard deviation of the known spectra instead of 2x, but then the risk of false exclusions increases.
It is difficult to know if a working analyst would make the same spectral exclusion and inclusion decisions as the researcher in this case. The issue with implementing the protocol spectral comparison suggested by SWGMAT is that enforcement is optional. Analysts are able to use the suggestions as a guideline but are not required to use any or all of the rules listed. It is strongly recommended that the forensic science community move away from these habits and find a more standardized method for conducting fiber comparisons.

### 4.2.3 Evaluation of Visual Comparison Method Reliability

The fiber SS and DS fiber comparisons described above are not considered statistically valid because the decisions for spectral inclusions and exclusions are completely reliant on the discretion of the analyst and subject to observer bias. There is no objective figure of merit used to make decisions for fiber comparisons. However, reliability can still be evaluated for this particular method using the percent accuracy. For SS and DS comparisons, the ground truth is known in every case. The percent accuracy is the sum of true spectral inclusions and true spectral exclusions (correct decisions) divided by the total number of all fiber comparisons made by visual analysis:

$$\% \text{ Accuracy} = \frac{\# \text{ True Correct SS Decisions} + \# \text{ True Correct DS Decisions}}{\text{Total \# Comparisons}} \times 100 \quad (4.1)$$

Table 4.7 shows the percent accuracy rate for the visual comparison method:

<table>
<thead>
<tr>
<th>Sum SS and DS Comparisons</th>
<th>Sum Correct Decisions</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>597</td>
<td>374</td>
<td>62.7</td>
</tr>
</tbody>
</table>

Table 4.7: Percent accuracy of visual comparison method for all comparisons
These results show that the visual comparison method is only 62.7% accurate (gives the correct decision 62.7% of the time) when applied to fiber comparisons. Note that the percentage might increase if a train fiber analyst were to replicate the spectral comparisons, however, this is a true testament of the weakness of this method. The visual comparison method may perform differently for different users.

4.3 Using Score Value to Characterize Fiber Comparisons

As described in section 3.6, the score value was used to characterize group membership for a questioned fiber for fiber comparisons after PCA was performed. Similar to the way the comparisons were made by spectral comparison method previously described, SS and DS comparisons are executed. The spectra were normalization to a baseline of zero and mean centered prior to PCA. The weighting parameter, $\gamma$, was optimized by testing with values of 0, 0.5, and 1; discrimination results were summarized for each value. The term discrimination was used instead of spectral exclusions since no visual analysis is considered here and because this method utilized a hard cutoff value for decision making. All numeric results, including score distance, score distance cutoff, orthogonal distance, orthogonal distance cutoff, and score value, from both SS and DS comparisons can be found in Appendix B.

4.3.1 PCA Results

PCA was used to reduce the dimensionality of large spectral data sets, but it was also used to visualize the natural groupings of the data through the generation of scores
plots. The scores from the averaged spectra that were normalized and mean centered were plotted for each of the three sample sets.

Figure 4.13 displays the scores plots with 95% confidence interval ellipses for each dyed fabric swatch where each score represents the average spectrum collected from a single fiber.

![M1-M6 Scores Plot](image)

**Figure 4.13: Scores plot for M1-M6 swatches**

The scores representing the spectra collected from fibers originating from swatch pairs (M1-M2, M3-M4, M5-M6) cluster near each other. For example, M5 (yellow) and M6 (gray) scores not only clustered away from the rest of the data but also clustered close together. M3D1-M4D1 (green and light blue) and M3D2-M4D2 scores (dark blue and
pink) overlap significantly. For the M1-M2 swatch pair clusters (black and red), there is a clear distinction between each individual grouping, however, there is not much scatter within the groups themselves. It is interesting that Figure 4.13 shows overlap between M2 and M3-M4 scores even though neither the chemical structure of the dyes used to color these materials nor the spectra collected from the fibers are similar, but it is likely that the third dimension of scores separates these clusters. The clustering that is present is due to the fact that the dyes chosen for this study are chemically similar; these similarities ultimately manifest in the above score plot.

Figure 4.14 displays the scores plots with 95% confidence interval ellipses for each blue yarn where each score represents the average spectrum collected from a single fiber.
Note that the yarns made by the same manufacturer (F-J and G-I) do in fact cluster together. Since dye structure is one of the factors that dictates spectral shape and these scores are representative of absorption profiles, chemical similarity of the dyes used in the yarn pairs was supported by the clustering shown above. It is also possible that different ratios of the same dyes were used to dye the yarns marketed from the same manufacturer in order to create different shades of the same color. Yarn H (red squares) exhibited a different absorption profile compared to the rest of the yarns, as seen in section 4.1.2, and as a result, its scores did not cluster near any of the other four groups.

Figure 4.15 displays the scores plots with 95% confidence interval ellipses for each denim fabric where each score represents the average spectrum collected from a single fiber.
Figure 4.15: Scores plot for Den3A-Den15 denims

The third and final scores plot exhibited a significant amount of overlap, and as a result, individual clusters were much less defined. The scores from Den3A (red) and Den5 (dark blue) overlap significantly as well as scores from Den8 (pink) and Den15 (black). Wherever there was appreciable overlap among score clusters, the researcher anticipated difficulty in discriminating between fibers of those groups using the score value method.

4.3.2 SS Comparisons

For the SS comparisons performed using the score value and the corresponding cutoff value to decide on discriminations between questioned and known fibers, the ground truth was that both the questioned and known fiber originated from the same swatch, yarn, or denim fabric. These types of comparisons are expected to result in score values less than the cutoff so that the questioned fiber does not discriminate from the known fibers. If any score values were greater than the cutoff value, the questioned fiber is ultimately discriminated from the knowns and was considered a false discrimination that contributed to Type I error.

4.3.2.1 Dyed Fabric Swatches

Table 4.8 shows a summary of the false discrimination results for SS comparisons among the dyed swatches.
Table 4.8: Summary of false discriminations for M1-M6

<table>
<thead>
<tr>
<th>Samples</th>
<th>M1</th>
<th>M2</th>
<th>M3D1</th>
<th>M3D2</th>
<th>M4D1</th>
<th>M4D2</th>
<th>M5</th>
<th>M6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Comparisons</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Number False Discriminations</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>γ = 0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>γ = 0.5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>γ = 1</td>
<td>80</td>
<td>7.5</td>
<td>18.8</td>
<td>38.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The best results (least amount of false discriminations) were obtained when the γ parameter is equal to zero, when the weight of the decision is placed entirely on the score distance (see equation 2.4). There are a total of six false discrimination results for the following comparisons: M2A3D1T1F1 v M2 Knowns, M2A3D2T3F1 v M2 Knowns, M3A1D1T1F1 v M3D1 Knowns, M3A1D1T2F2 v M3D1 Knowns, M5A5D2T3F1 v M5 Knowns, and M6A5D2T2F1 v M6 Knowns. Out of these six, the comparisons that were spectrally excluded as well are M2A3D2T3F1 v M2 Knowns and M5A5D2T3F1 v M5 Knowns. There were, however, some instances of comparisons that falsely discriminate using the score value but were determined to be spectral inclusions via the method employing visual comparisons. These included M2A3D1T1F1 v M2 Knowns, M3A1D1T1F1 v M3D1 Knowns, M3A1D1T2F2 v M3D1 Knowns, and M6A5D2T2F1 v M6 Knowns. Although the results from the score value method do not exactly match those of the visual comparison method, the method discussed here is completely objective and conducted without subjecting the fibers to observer bias. Lastly, there were some comparisons that were correct non-discriminations using the score value but were falsely excluded when spectral profiles were compared:
M1A5D2T1F1 v M1 Knowns, M1A5D2T2F4 v M1 Knowns, and M3A3D2T3F4 v M3D2 Knowns.

The differences between comparisons that falsely discriminated and those that did not can be visualized in the following Coomans plot.

![Coomans plots for false discrimination and correct non-discrimination](image)

**Figure 4.16: Coomans plots for false discrimination and correct non-discrimination**

When \( \gamma = 0.5 \), the questioned object, represented by the red Q, must lie to the left of the SD cutoff (vertical dotted line) and beneath the OD cutoff (horizontal dotted line) in order to be considered a non-discrimination. A discrimination occurs only when the questioned object lies both to the right of the score distance cutoff and above the orthogonal distance cutoff. Depending on \( \gamma \), the questioned object may lie to the left of the SD cutoff and above the OD cutoff and still be considered a non-discrimination; the
same is true if the questioned object lies to the right of the SD cutoff and below the OD cutoff.

The rates of false exclusion (section 4.2.1.1) and false discrimination are 6.25% and 7.5%, respectively. Overall, there was not much difference between the performances of the visual comparison and score value methods when applied to same sample comparisons. Each method reported the correct conclusion when the other failed to do so for a few specific comparisons.

4.3.2.2 Blue Yarns

Table 4.9 summarizes the false discrimination results for SS comparisons among the blue yarns.

Table 4.9: Summary of false discriminations for F-J

<table>
<thead>
<tr>
<th>Samples</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Comparisons</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number False Discriminations</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1 γ=0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3 γ=0.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3 γ=1</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% False Discriminations</td>
<td>16.0 γ=0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>52.0 γ=0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.0 γ=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As with the M1-M6 SS comparisons, the least amount of false discriminations occurred when the optimization parameter was equal to zero. However, even at the optimal value, the rate of false discrimination was equal to 16.0%. The four SS comparisons that falsely discriminated were F1 v F Knowns, G4 v G Knowns, I3 v I Knowns, and J5 v J Knowns. Out of these four, the only comparison that was considered a false spectral
exclusion via the visual comparison method was F1 v F Knowns. The remaining three comparisons were considered spectral inclusions.

In this case, using the score value for fiber discrimination decisions (16.0% false discrimination rate) does not perform as well as the spectral comparison method (4.0% false exclusion rate). Looking at the Coomans plots in Figure 4.17 for the four false discriminations, the position of the questioned objective relative to the SD and OD cutoffs can be seen.
The red questioned objects above were clearly positioned in the upper right quadrant allowing them to be clearly discriminated from the known objects in the lower left.
quadrant. It seems that the score value method is too robust when applied to spectra collected from the blue yarns.

4.3.2.3 Denim Fabrics

Table 4.10 summarizes the false discrimination results for SS comparisons among the denim fabrics.

Table 4.10: Summary of false discriminations for Den3A-Den15

<table>
<thead>
<tr>
<th>Sample</th>
<th>Den3A</th>
<th>Den4</th>
<th>Den5</th>
<th>Den6</th>
<th>Den8</th>
<th>Den9</th>
<th>Den15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number SS Comparisons</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Number False Discriminations</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% False Discriminations</td>
<td>10.7</td>
<td>0</td>
<td>32.1</td>
<td>44.6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The γ value that allows for the least amount of false discriminations (10.7%) for SS denim comparisons is 0. The six discriminations occur for the following comparisons: Den3AA1F2 v Den3A Knowns, Den3AA4F2 v Den3A Knowns, Den5A1F2 v Den5 Knowns, Den5A2F1 v Den Knowns, Den6A4F2 v Den6 Knowns, and Den8A1F2 v Den8 Knowns. Based on the SS results from all three datasets, it appears that the best results (fewest discriminations) occur when all the weight of discrimination decision making is placed on the SD or when γ=0.
4.3.3 DS Comparisons

For the DS comparisons performed using the score value and the corresponding cutoff value to decide on discriminations between questioned and knowns fibers, the ground truth was that they were sampled from different swatches, yarns, or denims fabrics. DS comparisons were expected to result in a score value that was greater than the cutoff so that the questioned fiber would discriminate from the known fibers. Any comparisons that failed to discriminate (false associations) were considered false non-discriminations that contributed to Type II error.

4.3.3.1 Dyed Fabric Swatches

Table 4.11 displays the false non-discrimination results summary for DS comparisons among the swatch pairs.

**Table 4.11: Summary of false associations for M1-M6 pairs**

<table>
<thead>
<tr>
<th>Sample Comparisons</th>
<th>M1vM2</th>
<th>M2vM1</th>
<th>M3D1v</th>
<th>M4D1v</th>
<th>M4D2v</th>
<th>M3D2v</th>
<th>M4D1</th>
<th>M3D1</th>
<th>M4D2</th>
<th>M3D2</th>
<th>M5vM6</th>
<th>M6vM5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Comparisons</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Number False Associations</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% False Associations</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The false association rate is 0% which supports the ground truth exactly. The spectral comparison method also performs with a 0% false inclusion rate for swatch DS
comparisons. The two methods perform equally as well as one another when applied to DS swatch comparisons. The fact that the score value results for M1-M6 swatches were consistent for all values of $\gamma$ implies that the swatches in each pair are significantly different from one another despite being dyed with dyes that were practically identical. Whichever differences were present in the corresponding absorption spectra are significant enough to allow for discrimination 100% of the time.

4.3.3.2 Blue Yarns

Table 4.12 summarizes the false association results for DS comparisons among the yarn pairs.

Table 4.12: Summary of false non-discriminations for F-J Pairs

<table>
<thead>
<tr>
<th>Sample Comparisons</th>
<th>F v J</th>
<th>J v F</th>
<th>G v I</th>
<th>I v G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Comparisons</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number False Association</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4 $\gamma=0$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2    $\gamma=0.5$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0    $\gamma=1$</td>
</tr>
<tr>
<td>Total Number Comparisons</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% False Association Rate</td>
<td>90.0 $\gamma=0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.0 $\gamma=0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 $\gamma=1$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When all weight is given to the OD, optimization parameter equal to 1, the rate of false associations is 0%; this supports the ground truth perfectly. When $\gamma=0$, the false association rate (90.0%) is only 5% greater than the false inclusion rate (85.0%) when the spectral comparisons method is employed for yarn DS comparisons, meaning that the two methods perform equally as poor as one another.
4.3.3.3 Denim Fabrics

Table 4.13 summarizes the false association results for DS comparisons among the denim pairs.
Table 4.13: Summary of false non-discriminations for Den3A-Den15

<table>
<thead>
<tr>
<th>Sample Comparisons</th>
<th>Number Comparisons</th>
<th>Number False Associations</th>
<th>Total Number Comparisons</th>
<th>% False Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>γ=0</td>
<td>γ=0.5</td>
<td>γ=1</td>
</tr>
<tr>
<td>Den4vDen3A</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den5vDen3A</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Den6vDen3A</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Den8vDen3A</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den9vDen3A</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Den15vDen3A</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den3AvDen4</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den5vDen4</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den6vDen4</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Den8vDen4</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Den9vDen4</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den15vDen4</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Den3AvDen5</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Den4vDen5</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den6vDen5</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Den8vDen5</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den9vDen5</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Den15vDen5</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Den3AvDen6</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Den4vDen6</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den5vDen6</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Den8vDen6</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Den9vDen6</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Den15vDen6</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Den3AvDen8</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den4vDen8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den5vDen8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den6vDen8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Den9vDen8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den15vDen8</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Den3AvDen9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Den4vDen9</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den5vDen9</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Den6vDen9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Den8vDen9</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den15vDen9</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den3AvDen15</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den4vDen15</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Den5vDen15</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Den6vDen15</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Den8vDen15</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Den9vDen15</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

115
The fewest false associations (23.5%) occur when $\gamma=1$, or when all the weight is placed on the OD in the score value calculation. DS comparisons for both the blue yarn and denim fabric data sets performed the best under the same conditions ($\gamma=1$), whereas the results from the dyed swatches do not change regardless of what $\gamma$ is equal to.

**4.3.4 Evaluation of Score Value Method Reliability**

The SS and DS comparisons are performed separately for each of the data sets because the ground truth is known, however, a working fiber analyst has no prior knowledge of the evidence before encountering it in the crime laboratory. In order to develop a universal method applied to conducting forensic fiber comparisons that is free of observer bias, an optimal $\gamma$ value must be agreed upon regardless of the ground truth. This way the fiber expert can execute any fiber comparison at any given time.

The percent accuracy was measured and compared to determine an optimal $\gamma$ value to perform any fiber comparisons using the score value method. Table 4.14 displays the results for method reliability for the score value method for all three data sets combined.

<table>
<thead>
<tr>
<th>Total Number Comparisons</th>
<th>597</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Correct Decisions</td>
<td></td>
</tr>
<tr>
<td>$\gamma=0$</td>
<td>390</td>
</tr>
<tr>
<td>$\gamma=0.5$</td>
<td>453</td>
</tr>
<tr>
<td>$\gamma=1$</td>
<td>443</td>
</tr>
<tr>
<td>% Accuracy</td>
<td></td>
</tr>
<tr>
<td>$\gamma=0$</td>
<td>65.3</td>
</tr>
<tr>
<td>$\gamma=0.5$</td>
<td>75.9</td>
</tr>
<tr>
<td>$\gamma=1$</td>
<td>74.2</td>
</tr>
</tbody>
</table>

Table 4.14: Percent accuracy of score value method for all comparisons
According to the percent accuracy values shown above, the best overall results are expected when $\gamma=0.5$; in this case, equal weight is given to both the SD and OD. The optimal value for SS comparisons only is $\gamma=0$ while the optimal value for DS comparisons only is $\gamma=1$. It is not surprising that the $\gamma$ value for all comparisons collectively is in between the two. The optimal percent accuracy at $\gamma=0.5$ is still only 75.9%, which means that the correct decision for a fiber comparison can only be expected 75.9% of the time. These results include the decisions made for the denim samples which undoubtedly caused a decrease in accuracy. As evidence from the amount of inherent spectral variability, the similarity in spectral profiles, and the degree of cluster overlap seen in the scores plots for the denim fibers, the researcher expected the score value method to perform poorly on the denim fabrics. If the denim results are removed from the calculation of percent accuracy, the value increases from 75.9% (all samples) to 84.9% (swatches and yarns only).

Although the score value method is able to eliminate observer bias by allowing fiber discrimination decisions to be based on a score cutoff value, the method reliability is lower than expected. The forensic science community requires a more accurate method in order to implement it toward casework. At 84.9% accuracy (disregarding denim fiber comparisons), it might be difficult for forensic experts to confidently testify to evidence generated by the score value method described above. It was encouraging, however, to see that when compared to the visual analysis method that only achieves 62.7% accuracy, the score value method performs better than it. In fact, when the
denim samples are not considered for evaluation of method reliability, the score value method is able to report a correct decision approximately 22% more often than the visual analysis method. Lastly, the score value method does not offer any support regarding the value of individual fiber comparisons. The NAS Report was very clear in demanding the creation of a way to evaluate the value of fiber evidence in addition to method reliability, but the score value method was unable to do both.

4.4 Using LRs to Characterize Fiber Comparisons

By understand the shortcomings of the previously described methods, the researcher were able to move forward by implementing a different way of conducting fiber comparisons using the LR. Similarities between spectra were evaluated using correlation coefficients (scores), probability density functions were modelled from those similarities, and finally the LRs calculated from those density functions were used to assign support to individual comparisons. A single comparison between a pair of fibers is assigned support to either of the competing hypotheses, $H_p$ or $H_d$, using the Evett verbal scale (Table 2.2).

The three datasets were analyzed differently than previously described and were divided into two new groups for this part of the study. The first new group contained the visible absorption spectra collected from both the dyed swatches and the blue yarns, while the second contained spectra from only the denim fabrics. The decision to combine the swatch and yarn spectra into one pool was made in order to create a more
comprehensive dataset from which the population model (probability density functions) were built. Note that the resulting dataset was neither representative for all existing fibers nor was it representative of the transferrable fiber population. The denim absorption spectra were treated separately in an attempt to better understand the complexity of these types of comparisons and to assign support to an individual denim cotton fiber comparisons. As discussed before, fiber comparisons of this type are extremely difficult to perform because denims are similarly dyed depending on what wash is desired. However, denim fibers are extremely common, so much so that many experts believe these comparisons have little to offer criminal investigations and cannot possess much value due to their great abundance. By treating the denim dataset separately, the researcher was able to explore this issue more thoroughly. Therefore, normal probability density functions for denim fiber comparisons were modeled separately from the rest of the samples for calculation of the LRs.

4.4.1 Swatches and Yarns

All absorption spectra were pretreated prior to conducting pairwise comparisons. Two iterations of testing was done so that the spectral data was first normalized to a baseline of zero and mean centered, and then it was normalized to between zero and one and mean centered. Optimization was done to determine which score allowed for highest percentage of correctly supported comparisons. The figure of merit used to identify the optimal coefficient is percent accuracy (Equation 4.1), where the numerator is the sum of the number of SS comparisons in support of $H_P$ and the number of DS comparisons
in support of $H_d$. Recall from section 3.7 that both SS and DS comparisons are performed for this method, and 80% of the resulting scores are used to create the training dataset while the remaining 20% of the scores are assigned to the test dataset. Based on ground truth, all SS comparisons should be assigned support in favor of $H_p$ whereas all of the DS comparisons should be assigned support in favor of $H_d$. Table 4.15 displays the summarized results from the optimization test.

Table 4.15: Percent accuracy results for swatches and yarns

<table>
<thead>
<tr>
<th>Normalization Method</th>
<th>Score</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>pearsonr</td>
<td>92.9</td>
</tr>
<tr>
<td>0</td>
<td>fisher-r</td>
<td>92.9</td>
</tr>
<tr>
<td>0</td>
<td>spearmanr</td>
<td>92.9</td>
</tr>
<tr>
<td>0</td>
<td>fisher-sr</td>
<td>92.9</td>
</tr>
<tr>
<td>0</td>
<td>cos_theta</td>
<td>89.4</td>
</tr>
<tr>
<td>01</td>
<td>pearsonr</td>
<td>97.7</td>
</tr>
<tr>
<td>01</td>
<td>fisher-r</td>
<td>97.7</td>
</tr>
<tr>
<td>01</td>
<td>spearmanr</td>
<td>96.5</td>
</tr>
<tr>
<td>01</td>
<td>fisher-sr</td>
<td>95.9</td>
</tr>
<tr>
<td>01</td>
<td>cos_theta</td>
<td>92.4</td>
</tr>
</tbody>
</table>

Overall, when using the 01 normalization method, the percent accuracy is higher than when normalizing the baseline to zero. The two scores that result in the highest accuracy rates are pearsonr and fisher-r. Both scores allow for 97.7% accuracy, but the distribution of the similarity values must be taken into account when determining which is best. Figure 4.18 displays two sets of histograms, one showing the distribution when the score is pearsonr and the other when the score is fisher-r.
In the histogram plot to the left (score: pearsonr), the DS distribution (red) is heavily skewed to the right, and the SS distribution (blue) skews drastically to the left; the distributions are in no way normal. There is, however, little overlap between the two distributions which ultimately lends itself to a more robust calculation of the LRs. If there is a large degree of overlap between the two distributions, then there will be greater room for confusion when assigning support to a fiber comparison whose score falls within the area of overlap. The corresponding probability density functions modelled after these distributions can be seen in Figure 4.19.
In the histogram plot on the right (Figure 4.18) both SS (blue) and DS (red) distributions show a much more normal distribution with little skewing to either side of the plot (SS skewed right and DS skewed left). However, a larger degree of overlap is present in both the score distributions and resulting probability density functions (right plot in Figure 4.19). Scores from SS and DS comparisons that exhibit a normal distribution and exhibit a very small amount of overlap allow for the most robust LRs to be calculated. Based on the percent accuracy rate and the shapes of the density function curves, the optimal score was chosen to be fisher-r. The confusion matrix below in Table 4.16 displays the predicted outcomes of the LR calculations versus what the actual tests results were.
Table 4.16: Compiled LR results for swatches and yarns

<table>
<thead>
<tr>
<th>Predicted Outcome</th>
<th>Actual Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_p$</td>
<td>77</td>
</tr>
<tr>
<td>$H_d$</td>
<td>3</td>
</tr>
</tbody>
</table>

According to these results, 77 of the SS comparisons in the test set were assigned support in favor of $H_p$, and 89 of the DS comparisons were assigned support in favor of $H_d$. The sum of these two values (166) was the numerator in the ratio used to calculate percent accuracy (97.7%). Out of the 170 SS and DS comparisons only four of those were incorrectly support in favor of the opposing hypothesis. The instances where comparisons were assigned support in favor of the incorrect hypothesis were investigated further. Table 4.17 shows a data table where the number of SS and DS comparisons that were assigned varying degrees of support according to Evett’s verbal scale can be seen.

Table 4.17: Expanded LR results for swatches and yarns

<table>
<thead>
<tr>
<th>Predicted Outcome</th>
<th>Support for $H_p$</th>
<th>Support for $H_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VS</td>
<td>S</td>
</tr>
<tr>
<td>$H_p$</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>$H_d$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Out of the 170 comparisons performed that make up the test dataset, 80 were truly SS comparisons and 90 were truly DS comparisons. A majority of the SS comparisons (61) were assigned either very strong, strong, or moderately strong support in favor of $H_p$, whereas most of the DS comparisons (70) were assigned to the same degrees of
support in favor of $H_d$. It is interesting that more SS comparisons were given limited support (10) than moderate support (6), but this is not the case for DS comparisons. Recall that the training dataset was only comprised of comparisons between dyed swatch and yarn pairs that were chosen specifically because the dyes exhibited highly similar visible absorption profiles. Because the scope was so limited for these tests, it would be beneficial to add many more textile samples in order to broaden the population dataset from which the density functions were modelled. Despite this, the results also show that the four comparisons that were incorrectly supported (one SS and three DS) were only assigned limited support in favor of the opposing hypothesis. Although these results were undesirable, it was reassuring that only the least amount of support possible as offered to the competing hypothesis. The LR method was still able to achieve a 97.7% accuracy rate when applied to swatches and yarns. Compared to the accuracy rate of the score value method at $\gamma=0.5$ (75.9% including denim comparisons, 84.9% excluding denim comparisons), the LR method not only outperformed the score value method in terms of reporting correct results but was also able to assign the support to an individual fiber comparison.

Furthermore, the validity of the LR method can be evaluated using the data in Table 4.16. The researcher calculated the ratio between the different probabilities (probability of assigning support to either hypothesis) given the ground truth of a comparison. The probability of assigning support to $H_p$ given that a comparison was truly a SS comparison divided by the probability of assigning support to $H_p$ given that a
comparison was truly a DS comparisons was given by \( \frac{\Pr(\text{assigning support to } H_p|SS)}{\Pr(\text{assigning support to } H_p|DS)} = \frac{77/80}{1/90} = 86.6 \). This meant that the LR method was 86.6 times more likely to offer support to \( H_p \) when a comparison was truly between fibers from a common source than to arrive at the same conclusions when the fiber do not actually come from the same source. Conversely, the probability of assigning support to \( H_d \) given that a comparison was truly a DS comparison divided by the probability of assigning support to \( H_d \) given that a comparison was actually a SS comparison was given by \( \frac{\Pr(\text{assigning support to } H_d|DS)}{\Pr(\text{assigning support to } H_d|SS)} = \frac{89/90}{3/80} = 26.4 \). This meant that the LR method was 26.4 times more likely to offer support to \( H_d \) when a comparison was truly between fibers that originated from different sources than to arrive at the same conclusion when the fibers actually do come from the same source. For this particular dataset (swatches-yarns), conclusions derived from the LR method that correctly support \( H_p \) are more meaningful than conclusions that correctly support \( H_d \). That is, the evidential value or the weight of the evidence is greater for a decision in support of \( H_p \) than \( H_d \).

### 4.4.2 Denim Fabrics

The denim spectra were treated in the same way as described in section 4.4.1 prior to optimization testing. Results from optimization tests are displayed in Table 4.18.
Table 4.18: Percent accuracy results for denims

<table>
<thead>
<tr>
<th>Normalization Method</th>
<th>Score</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>pearsonr</td>
<td>64.2</td>
</tr>
<tr>
<td>0</td>
<td>fisher-r</td>
<td>62.9</td>
</tr>
<tr>
<td>0</td>
<td>spearmanr</td>
<td>63.5</td>
</tr>
<tr>
<td>0</td>
<td>fisher-sr</td>
<td>63.2</td>
</tr>
<tr>
<td>0</td>
<td>cos_theta</td>
<td>56.4</td>
</tr>
<tr>
<td>01</td>
<td>pearsonr</td>
<td>62.5</td>
</tr>
<tr>
<td>01</td>
<td>fisher-r</td>
<td>67.1</td>
</tr>
<tr>
<td>01</td>
<td>spearmanr</td>
<td>63.5</td>
</tr>
<tr>
<td>01</td>
<td>fisher-sr</td>
<td>65.5</td>
</tr>
<tr>
<td>01</td>
<td>cos_theta</td>
<td>60.9</td>
</tr>
</tbody>
</table>

Again, the highest percent accuracy occurred when the data was normalized from zero to one and when the fisher-r was the score. Table 4.19 shows the predicted outcomes of the LR denim calculations versus what the actual tests results were.

Table 4.19: Compiled LR results for denims

<table>
<thead>
<tr>
<th>Actual Test Results</th>
<th>$H_P$</th>
<th>$H_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Outcome</td>
<td>$H_P$</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>$H_d$</td>
<td>91</td>
</tr>
</tbody>
</table>

This results summary shows that out of the 39 SS comparisons represented in the test set, 29 of those were correctly supported in favor of $H_P$ while 10 were incorrectly supported in favor of $H_d$. Out of the 268 DS comparisons, 177 of those were correctly supported in favor of $H_d$ while 91 (a rather large amount) were incorrectly supported in favor of $H_P$. For denims, there was a significantly greater number of comparisons that
were incorrectly supported compared to the swatch and yarn results. This is not surprising, however, because the LRs were calculated from probability density functions that exhibited a large degree of overlap. Recall that the scores evaluate the similarity or correlation between a pair of spectra; if the absorption spectra are highly similar then there will not be a wide spread in the score distribution, and as a result, there will be a larger overlap in the probability curves. Figure 4.20 compares two histograms showing the distribution of scores from the swatch-yarn comparisons and the denim comparisons.

![Figure 4.20: Swatch-yarn score distributions vs denim score distributions](image)

The histogram on the left shows two distinct distributions of scores for both swatch and yarn comparisons (SS distribution shown in blue and DS distribution shown in red) with some overlap between them. The histogram on the right (scores resulting from denim
comparisons) shows that the SS scores fall nearly completely within the distribution of the DS scores. The implications of this were that it would be very difficult to correctly support a true SS comparison because most of the scores are among those for DS comparisons. Similarly for the true DS comparisons, these have a greater chance of being incorrectly supported in favor of $H_p$ because of the high degree of score distribution overlap.

Table 4.20 contains the expanded LR results showing what degrees of support are offered to the SS and DS comparisons.

Table 4.20: Expanded LR results for denims

<table>
<thead>
<tr>
<th>Predicted Outcome</th>
<th>Support for $H_p$</th>
<th>Support for $H_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>M</td>
</tr>
<tr>
<td>$H_p$</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>$H_d$</td>
<td>91</td>
<td>127</td>
</tr>
</tbody>
</table>

These results prove that correct assignment of support is difficult to obtain when there is such a large overlap in the score distributions. The 29 SS comparisons that are correctly supported are only offered limited support whereas the 10 incorrectly supported comparisons are offered moderate and moderately strong support. Just less than half (127) of the DS comparisons are only given limited support in favor of $H_d$, and nearly a third (91) DS comparisons are incorrectly assigned support in favor of $H_p$. Overall, these results offer empirical evidence to the long standing belief that denim cotton fibers pose a serious challenge to forensic examiners. Similar to the suggestion made for the swatch and yarn experiments, additional denims of different colors and
wash styles should be added to the dataset in the future in order to create a more complete population set. Because weak evidential value was assigned to individual denim fiber comparisons, these results may suggest that the LR method is simply not appropriate for denim fabrics. Further work is needed to develop a unique approach to these types of forensic comparisons.

Validity of the LR method applied to denim comparisons was evaluated in the same way as described previously for swatch and yarn comparisons. The probability of assigning support to \( H_p \) given that a comparison is truly a SS comparison divided by the probability of assigning support to \( H_p \) given that a comparison is truly a DS comparison is given by:

\[
\frac{P(\text{assigning support to } H_p|\text{SS})}{P(\text{assigning support to } H_p|\text{DS})} = \frac{29/39}{91/268} = 2.19.
\]

This meant that the method was 2.19 times more likely to offer support in favor of \( H_p \) when a comparison is truly of the SS variety than to arrive at the same conclusion for a truly DS comparison. Conversely, the probability of assigning support to \( H_d \) given that a comparison is truly a DS comparison divided by the probability of assigning support to \( H_d \) given that a comparison is truly a SS comparison is given by:

\[
\frac{P(\text{assigning support to } H_d|\text{DS})}{P(\text{assigning support to } H_d|\text{SS})} = \frac{177/268}{10/39} = 2.58.
\]

In this case, the method was 2.58 times more likely to offer support to \( H_d \) when the fibers are actually from different sources than to arrive at the same conclusions for a truly SS comparison. Approximately, the same degree of evidential weight can be expected for both types of assignments using the LR method applied to denim fabrics.
In summary, the visual comparison method of fiber analysis performed with an overall accuracy rate of 62.7% when applied to fibers from all three datasets. The percent accuracy was believed to be changeable based on the opinion of the fiber analyst and his training and experience. The subjectivity of the visual comparison method was and continues to be undesirable. The score value method performed with 75.9% accuracy at $\gamma=0.5$ when applied to fibers from all three datasets. Because denim comparisons were known to be difficult and in order to evaluate the reliability of the score value method without the influence of the denim comparisons, the percent accuracy was calculated to be 84.8% for swatch and yarn comparisons only. These results showed an improvement in method performance compared to the visual comparison method, but a higher accuracy rate was still desired. Furthermore, evidential value of individual fiber comparisons was not able to be reported using the score value method. When applied to swatch and yarn comparisons only, the LR method performed with a 97.7% accuracy rate (fisher-r was used as the score metric and spectral data was normalized from zero to one). The LR by far outperformed both the visual comparison and score value methods and was able to assign strong, correct support to either $H_p$ or $H_d$ for an individual fiber comparison. When the LR method was applied to denim comparisons (same score metric and normalization for swatch and yarn comparisons), the percent accurate dropped to 67.1%. Additionally, weak, correct support was assigned to $H_p$ and $H_d$ for SS and DS comparisons, respectively.
CHAPTER 5: CONCLUSIONS

The goal of this research was to apply chemometric analysis to forensic fiber comparisons in order to develop a completely objective method of comparing visible absorption spectra collected from textile fibers. Literature reports that in some instances where fibers are highly similar in color, traditional methods of visual comparisons of spectral profiles can fail to report the correct results or give rise to inconclusive decisions made by the fiber analyst. For this reason, samples that were dyed similarly to one another were chosen for this research.

Fiber comparison were conducting using the visual comparison method in order to mimic those being made in crime laboratories in the United States. Although other fiber characteristics besides visible absorption spectra are considered when performing fiber examinations, the comparison of fiber color using MSP is rapid. The MSP is not capable of harnessing characteristic spectral features that can lead to dye identification or classification, the aim of many forensic comparisons is to characterize evidence (often through pattern recognition), not to identify its chemical constituents.

Based on the results from investigation of inherent spectral variability and conducting fiber comparisons using the visual comparison method, it was confirmed that the application of chemometric techniques could be very helpful in developing a method for comparing similarly colored textile fibers. The application of PCA in the score value method helped extract latent information in the spectral data that was unutilized for
performing forensic fiber examinations. Implementing the score value as the metric used to decide if fiber comparisons resulted in discriminations or associations allowed the researcher to perform fiber analysis in a way that was free from observer bias. While the score value method performed better than the visual comparison method in terms of percent accuracy, a universal method that could be applied to any fiber comparison regardless of the ground truth and that could report the evidential value of an individual fiber comparison was still needed.

Using LRs to evaluate forensic fiber comparisons is different than the visual comparison and score value methods because background population data was needed to model probability density functions. Also, SS and DS comparisons were treated and analyzed collectively whereas in the other procedures it was required that these comparisons be performed separately. Most importantly, the LR method was able to assign varying degrees of support to individual fiber comparisons while also performing fiber analysis that was free from observer bias. When applied to swatch and yarn comparisons only, the LR method was able to achieve the highest percent accuracy rate of all methods described in this work.

In the future, the LR method can be improved by sampling a much larger collection of textile fibers. The larger the population dataset is, the more representative it is considered. In addition to custom dyed fabric swatches and yarns, spectra should be collected from other types of fabrics (carpets, furniture, etc.) including various pieces of
clothing (shirts, socks, etc.). Harnessing and organizing large collections of spectra may require the development of a spectral database or collaboration efforts with institutions that have already begun work in creating fiber databases. It is also recommended that collaborations with experts that specialize in the application of Bayesian methods to forensic science be explored in an attempt to fortify the LR method described here. There is still much more work that needs to be done with regards to method validation in order to develop an objective method of analysis for fiber comparisons that can be applied to casework. With regard to denim fibers, more exploration is needed to identify a method of analysis that is robust enough to handle forensic comparisons of this type. Currently, the reality is that denim fibers are difficult to discriminate from one another, and the fiber expert must use extreme caution when reporting findings in the final forensic report and when testifying in court.
APPENDIX A:
VISUAL COMPARISON OF ABSORPTION SPECTRA RESULT PLOTS
Figure 5.1: M1-M6 SS spectral comparison results
151
Figure 5.2: F-J SS spectral comparison results
Den3AA1F1 vs Known Spectra

Den3AA1F2 vs Known Spectra

Den3AA2F1 vs Known Spectra

Den3AA2F2 vs Known Spectra

Den3AA3F1 vs Known Spectra

Den3AA3F2 vs Known Spectra
Figure 5.3: Den3A-Den15 SS spectral comparison results
Figure 5.4: M1-M6 DS spectral comparisons results
Figure 5.5: F-J and G-I DS spectral comparisons results
Figure 5.6: Den3A-Den15 DS spectral comparisons results
APPENDIX B:
NUMERIC RESULTS FROM COMPARISONS USING SCORE VALUE
```r
source("C:\Research\PhyS Evidence\Personal\Allie Flores\Code\R files\Mahalanobis outliers\SD-0D SS v6.r")

# x-read.table("M1 spectra.csv", sep="", header=T)

> sddd(x, qtile=0.975, nm = "0", cent="mc", npci=-.95, OR = "", alphaMCd=0.75, sdddgamma=0)

```
```r
> x <- read.table("M101 spectra.csv", sep="", header=TRUE)
> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5000gamma=0)

<table>
<thead>
<tr>
<th>SampleID</th>
<th>Mahalanobisd</th>
<th>NPC</th>
<th>Score_Dist</th>
<th>SD_Dist</th>
<th>Blank1</th>
<th>Orthogonal_Dist</th>
<th>OD_Dist</th>
<th>Odds</th>
<th>Blank</th>
<th>Score</th>
<th>Value</th>
<th>SDOO_Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>M193217F1</td>
<td>3.2154281</td>
<td>2.712603</td>
<td>3.2154281</td>
<td>T</td>
<td>0.0520725</td>
<td>0.0702926</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F2</td>
<td>2.9490973</td>
<td>2.712603</td>
<td>2.9490973</td>
<td>T</td>
<td>0.0470600</td>
<td>0.0738016</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F3</td>
<td>1.7614517</td>
<td>2.712603</td>
<td>1.7614517</td>
<td>T</td>
<td>0.0481375</td>
<td>0.0707271</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F4</td>
<td>0.9661363</td>
<td>2.712603</td>
<td>0.9661363</td>
<td>T</td>
<td>0.0657292</td>
<td>0.0886699</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F5</td>
<td>1.2173797</td>
<td>2.712603</td>
<td>1.2173797</td>
<td>T</td>
<td>0.0500721</td>
<td>0.0675386</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F6</td>
<td>0.7623879</td>
<td>2.712603</td>
<td>0.7623879</td>
<td>T</td>
<td>0.0654333</td>
<td>0.0621981</td>
<td>T</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F7</td>
<td>0.4470802</td>
<td>2.712603</td>
<td>0.4470802</td>
<td>T</td>
<td>0.0458372</td>
<td>0.0605739</td>
<td>T</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F8</td>
<td>0.6138357</td>
<td>2.712603</td>
<td>0.6138357</td>
<td>T</td>
<td>0.0384427</td>
<td>0.0703018</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F9</td>
<td>0.8698777</td>
<td>2.712603</td>
<td>0.8698777</td>
<td>T</td>
<td>0.0479420</td>
<td>0.0769847</td>
<td>F</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M193217F10</td>
<td>2.4777927</td>
<td>2.712603</td>
<td>2.4777927</td>
<td>T</td>
<td>0.0883236</td>
<td>0.0659452</td>
<td>T</td>
<td>0.91074</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5000gamma=5)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5005gamma=1)

> x <- read.table("M102 spectra.csv", sep="", header=TRUE)
> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5000gamma=0)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5005gamma=5)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5005gamma=1)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5000gamma=5)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5005gamma=1)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5000gamma=0)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5005gamma=5)

> SDOO(x, qttile=0.975, nm = "oc", nc = "nc", ncp = 95, or = "", alphaco.75, 5005gamma=1)

```

192
```r
# Read the CSV file
x <- read.table("M401_spectra.csv", sep="", header=T)

# Display the subset of data
sdoo(x, qtile=0.975, nm="0", cent="mc", npi=-.95, or="", alphancd=0.75, sdooalpha=0)

# Display the subset of data
sdoo(x, qtile=0.975, nm="0", cent="mc", npi=-.95, or="", alphancd=0.75, sdooalpha=5)

# Display the subset of data
sdoo(x, qtile=0.975, nm="0", cent="mc", npi=-.95, or="", alphancd=0.75, sdooalpha=0)

# Display the subset of data
sdoo(x, qtile=0.975, nm="0", cent="mc", npi=-.95, or="", alphancd=0.75, sdooalpha=5)
```
Figure 5.7: M1-M6 SS score value results
```r
> x <- read.table("1-5 spectra.csv", sep="", header=T)
> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=0)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 5.0024892 2.716203 5.0624892 T 0.3514030 0.1015485 T 0.978072 T
2 2 0.5044936 2.716203 0.5044936 T 0.1002114 0.1460027 F 0.978072 T
3 3 1.7634531 2.716203 1.7634531 F 0.2178587 0.0924477 F 0.978072 T
4 4 0.6234037 2.716203 0.6234037 F 0.1313841 0.1031969 F 0.978072 T
5 5 2.5451901 2.716203 2.5451901 F 0.2610066 0.2721126 T 0.978072 T

> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=3)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 5.0624892 2.716203 5.0624892 T 0.3514030 0.1018545 T 5.264473 T
2 2 0.5044936 2.716203 0.5044936 T 0.1002114 0.1460027 F 5.264473 T
3 3 1.7634531 2.716203 1.7634531 F 0.2178587 0.0920477 T 5.264473 T
4 4 0.6234037 2.716203 0.6234037 F 0.1313841 0.1031969 F 5.264473 T
5 5 2.5451901 2.716203 2.5451901 F 0.2610066 0.2721126 T 5.264473 T

> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=1)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 5.0624892 2.716203 5.0624892 T 0.3514030 0.1018545 T 9.591859 T
2 2 0.5044936 2.716203 0.5044936 T 0.1002114 0.1460027 F 9.591859 T
3 3 1.7634531 2.716203 1.7634531 F 0.2178587 0.0920477 T 9.591859 T
4 4 0.6234037 2.716203 0.6234037 F 0.1313841 0.1031969 F 9.591859 T
5 5 2.5451901 2.716203 2.5451901 F 0.2610066 0.2721126 T 9.591859 T

> x <- read.table("G 1-5 spectra.csv", sep="", header=T)
> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=0)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 0.8738816 2.716203 0.8738816 F 0.1232034 0.1371764 T 0.3458071 F
2 2 1.0763509 2.716203 1.0763509 T 0.0938926 0.1519910 F 0.3458071 F
3 3 1.6420927 2.716203 1.6420927 F 0.1571360 0.1186598 T 0.3458071 F
4 4 3.8911581 2.716203 3.8911581 F 0.4307065 0.0363435 T 0.3458071 F
5 5 0.9199282 2.716203 0.9199282 T 0.1226725 0.0841232 T 0.3458071 F

> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=3)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 0.8738816 2.716203 0.8738816 F 0.1232034 0.1371764 T 1.458248 T
2 2 1.0763509 2.716203 1.0763509 T 0.0938926 0.1519910 F 1.458248 T
3 3 1.6420927 2.716203 1.6420927 F 0.1571360 0.1186598 T 1.458248 T
4 4 3.8911581 2.716203 3.8911581 F 0.4307065 0.0363435 T 1.458248 T
5 5 0.9199282 2.716203 0.9199282 T 0.1226725 0.0841232 T 1.458248 T

> x <- read.table("m 1-5 spectra.csv", sep="", header=T)
> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=0)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 1.801072 2.716203 1.801072 F 0.2174799 0.1080620 T 0.375034 F
2 2 2.117478 2.716203 2.117478 F 0.2223320 0.1324686 F 0.375034 F
3 3 1.253284 2.716203 1.253284 F 0.2021980 0.0914849 F 0.375034 F
4 4 1.817437 2.716203 1.817437 F 0.1976768 0.1629358 F 0.375034 F
5 5 1.195052 2.716203 1.195052 F 0.1263103 0.0996793 F 0.375034 F

> sdod(x, qtile=0.975, nm = "0", cent="mc", rpci=95, OR = "", alphacc=0.75, sdodgamma=3)

> sampleID MahalanobisD NPC SDodCutoff Score_PDist SD_Dist blanki Orthogonal_Dist OD_cutoff ODDis blank SCare Value SDOD_DIST
1 1 1.801072 2.716203 1.801072 F 0.2174799 0.1080620 T 2.170771 T
2 2 2.117478 2.716203 2.117478 F 0.2223320 0.1324686 F 2.170771 T
3 3 1.253284 2.716203 1.253284 F 0.2021980 0.0914849 F 2.170771 T
4 4 1.817437 2.716203 1.817437 F 0.1976768 0.1629358 F 2.170771 T
5 5 1.195052 2.716203 1.195052 F 0.1263103 0.0996793 F 2.170771 T
```
Figure 5.8: F-J SS score value results
```r
# Read table
x=read.table("D:/Research PHYS Evidence/Personal/Allie Flores/Code/R files/Kahalanobis outliers/SD-OD SS_V6.r")

# SOD(x, qtile=0.975, cm = "0", center="nc", npci=.95, or = "~", alphacd=0.75, SODgamma=0)

```
```r
> source("C:\Research\PHS Evidence\Personal\Allie Flores\Code\R files\Mahalanobis outliers\SD-DD SS_v6.r")
> x=read.table("deni indiv.csv", sep="", header=T)

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Den1A1F</td>
<td>1.82980</td>
<td>2.716203</td>
<td>1.82980</td>
<td>F</td>
<td>0.183158</td>
<td>0.355187</td>
<td>F</td>
<td>0.390325</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Den1A1F</td>
<td>2.716689</td>
<td>2.716203</td>
<td>2.716689</td>
<td>T</td>
<td>0.459078</td>
<td>0.286834</td>
<td>T</td>
<td>0.390325</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Den1A2F</td>
<td>7.117367</td>
<td>2.716203</td>
<td>7.117367</td>
<td>T</td>
<td>0.951868</td>
<td>0.427264</td>
<td>T</td>
<td>0.390325</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Den1A2F</td>
<td>0.093680</td>
<td>2.716203</td>
<td>0.093680</td>
<td>F</td>
<td>0.249034</td>
<td>0.251564</td>
<td>F</td>
<td>0.390325</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Den1A3F</td>
<td>1.040073</td>
<td>2.716203</td>
<td>1.040073</td>
<td>T</td>
<td>0.249872</td>
<td>0.399121</td>
<td>T</td>
<td>0.390325</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Den1A3F</td>
<td>0.784359</td>
<td>2.716203</td>
<td>0.784359</td>
<td>T</td>
<td>0.409733</td>
<td>0.321117</td>
<td>T</td>
<td>0.390325</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Den1A4F</td>
<td>0.727347</td>
<td>2.716203</td>
<td>0.727347</td>
<td>T</td>
<td>0.145138</td>
<td>0.232382</td>
<td>T</td>
<td>0.390325</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Den1A4F</td>
<td>1.067536</td>
<td>2.716203</td>
<td>1.067536</td>
<td>T</td>
<td>0.436191</td>
<td>0.326406</td>
<td>T</td>
<td>0.390325</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Den2A1F</td>
<td>1.82980</td>
<td>2.716203</td>
<td>1.82980</td>
<td>F</td>
<td>0.183158</td>
<td>0.355187</td>
<td>F</td>
<td>1.336369</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Den2A1F</td>
<td>2.716689</td>
<td>2.716203</td>
<td>2.716689</td>
<td>T</td>
<td>0.459078</td>
<td>0.286834</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Den2A2F</td>
<td>7.117367</td>
<td>2.716203</td>
<td>7.117367</td>
<td>T</td>
<td>0.951868</td>
<td>0.427264</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Den2A2F</td>
<td>0.093680</td>
<td>2.716203</td>
<td>0.093680</td>
<td>F</td>
<td>0.249034</td>
<td>0.251564</td>
<td>F</td>
<td>1.336369</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Den2A3F</td>
<td>1.040073</td>
<td>2.716203</td>
<td>1.040073</td>
<td>T</td>
<td>0.249872</td>
<td>0.399121</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Den2A3F</td>
<td>0.784359</td>
<td>2.716203</td>
<td>0.784359</td>
<td>T</td>
<td>0.409733</td>
<td>0.321117</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Den2A4F</td>
<td>0.727347</td>
<td>2.716203</td>
<td>0.727347</td>
<td>T</td>
<td>0.145138</td>
<td>0.232382</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Den2A4F</td>
<td>1.067536</td>
<td>2.716203</td>
<td>1.067536</td>
<td>T</td>
<td>0.436191</td>
<td>0.326406</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Den3A1F</td>
<td>1.82980</td>
<td>2.716203</td>
<td>1.82980</td>
<td>F</td>
<td>0.183158</td>
<td>0.355187</td>
<td>F</td>
<td>1.336369</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Den3A1F</td>
<td>2.716689</td>
<td>2.716203</td>
<td>2.716689</td>
<td>T</td>
<td>0.459078</td>
<td>0.286834</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Den3A2F</td>
<td>7.117367</td>
<td>2.716203</td>
<td>7.117367</td>
<td>T</td>
<td>0.951868</td>
<td>0.427264</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Den3A2F</td>
<td>0.093680</td>
<td>2.716203</td>
<td>0.093680</td>
<td>F</td>
<td>0.249034</td>
<td>0.251564</td>
<td>F</td>
<td>1.336369</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Den3A3F</td>
<td>1.040073</td>
<td>2.716203</td>
<td>1.040073</td>
<td>T</td>
<td>0.249872</td>
<td>0.399121</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Den3A3F</td>
<td>0.784359</td>
<td>2.716203</td>
<td>0.784359</td>
<td>T</td>
<td>0.409733</td>
<td>0.321117</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Den3A4F</td>
<td>0.727347</td>
<td>2.716203</td>
<td>0.727347</td>
<td>T</td>
<td>0.145138</td>
<td>0.232382</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Den3A4F</td>
<td>1.067536</td>
<td>2.716203</td>
<td>1.067536</td>
<td>T</td>
<td>0.436191</td>
<td>0.326406</td>
<td>T</td>
<td>1.336369</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>
```

198
```r
> source("C:\\research PHYS evidence\\Personal\\Allie Flores\\Code\\R files\\Mahalanobis outliers SD-00 SS_v6.r")
> x=read.table("denf_indiv.csv", sep="", header=T)

<table>
<thead>
<tr>
<th>SampleID</th>
<th>MahalanobisD</th>
<th>NPC</th>
<th>Score</th>
<th>Dist</th>
<th>SD_D</th>
<th>blank</th>
<th>Orthogonal_Dist</th>
<th>OD_Cutoff</th>
<th>Odds</th>
<th>blank</th>
<th>OD_Cutoff</th>
<th>Score</th>
<th>Value</th>
<th>SD_D</th>
<th>blank</th>
<th>OD_Cutoff</th>
<th>Odds</th>
<th>blank</th>
<th>SD_D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.309246</td>
<td>2.716203</td>
<td>0.309246</td>
<td>F</td>
<td>201.7374</td>
<td>230.4613</td>
<td>F</td>
<td>0.0982641</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.3148975</td>
<td>2.716203</td>
<td>0.3148975</td>
<td>T</td>
<td>1.0993006</td>
<td>334.8036</td>
<td>T</td>
<td>0.0982641</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.688512</td>
<td>2.716203</td>
<td>0.858912</td>
<td>T</td>
<td>0.5019080</td>
<td>0.5792139</td>
<td>T</td>
<td>0.0982641</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.686783</td>
<td>2.716203</td>
<td>0.686783</td>
<td>T</td>
<td>0.5211077</td>
<td>0.2892390</td>
<td>T</td>
<td>0.0982641</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.5121312</td>
<td>2.716203</td>
<td>0.2512131</td>
<td>T</td>
<td>0.2921097</td>
<td>0.2805050</td>
<td>T</td>
<td>0.0982641</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.4209181</td>
<td>2.716203</td>
<td>0.4209181</td>
<td>T</td>
<td>0.3480695</td>
<td>0.2973605</td>
<td>T</td>
<td>0.0982641</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.034307</td>
<td>2.716203</td>
<td>0.034307</td>
<td>F</td>
<td>0.1703935</td>
<td>0.3493912</td>
<td>F</td>
<td>0.0982641</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.266072</td>
<td>2.716203</td>
<td>0.266072</td>
<td>F</td>
<td>0.4105325</td>
<td>0.3112040</td>
<td>T</td>
<td>0.0982641</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SampleID</th>
<th>MahalanobisD</th>
<th>NPC</th>
<th>Score</th>
<th>Dist</th>
<th>SD_D</th>
<th>blank</th>
<th>Orthogonal_Dist</th>
<th>OD_Cutoff</th>
<th>Odds</th>
<th>blank</th>
<th>OD_Cutoff</th>
<th>Score</th>
<th>Value</th>
<th>SD_D</th>
<th>blank</th>
<th>OD_Cutoff</th>
<th>Odds</th>
<th>blank</th>
<th>SD_D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.309246</td>
<td>2.716203</td>
<td>0.309246</td>
<td>F</td>
<td>0.4377374</td>
<td>0.3304613</td>
<td>F</td>
<td>1.319148</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.3148975</td>
<td>2.716203</td>
<td>0.3148975</td>
<td>T</td>
<td>1.0993006</td>
<td>334.8036</td>
<td>T</td>
<td>1.319148</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.688512</td>
<td>2.716203</td>
<td>0.858912</td>
<td>T</td>
<td>0.5019080</td>
<td>0.5792139</td>
<td>T</td>
<td>1.319148</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.686783</td>
<td>2.716203</td>
<td>0.686783</td>
<td>T</td>
<td>0.5211077</td>
<td>0.2892390</td>
<td>T</td>
<td>1.319148</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.5121312</td>
<td>2.716203</td>
<td>0.2512131</td>
<td>T</td>
<td>0.2921097</td>
<td>0.2805050</td>
<td>T</td>
<td>1.319148</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.4209181</td>
<td>2.716203</td>
<td>0.4209181</td>
<td>T</td>
<td>0.3480695</td>
<td>0.2973605</td>
<td>T</td>
<td>1.319148</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.034307</td>
<td>2.716203</td>
<td>0.034307</td>
<td>F</td>
<td>0.1703935</td>
<td>0.3493912</td>
<td>F</td>
<td>1.319148</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.266072</td>
<td>2.716203</td>
<td>0.266072</td>
<td>F</td>
<td>0.4105325</td>
<td>0.3112040</td>
<td>T</td>
<td>1.319148</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SampleID</th>
<th>MahalanobisD</th>
<th>NPC</th>
<th>Score</th>
<th>Dist</th>
<th>SD_D</th>
<th>blank</th>
<th>Orthogonal_Dist</th>
<th>OD_Cutoff</th>
<th>Odds</th>
<th>blank</th>
<th>OD_Cutoff</th>
<th>Score</th>
<th>Value</th>
<th>SD_D</th>
<th>blank</th>
<th>OD_Cutoff</th>
<th>Odds</th>
<th>blank</th>
<th>SD_D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.309246</td>
<td>2.716203</td>
<td>0.309246</td>
<td>F</td>
<td>0.4377374</td>
<td>0.3304613</td>
<td>F</td>
<td>0.2660728</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.3148975</td>
<td>2.716203</td>
<td>0.3148975</td>
<td>T</td>
<td>1.0993006</td>
<td>334.8036</td>
<td>T</td>
<td>0.2660728</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.688512</td>
<td>2.716203</td>
<td>0.858912</td>
<td>T</td>
<td>0.5019080</td>
<td>0.5792139</td>
<td>T</td>
<td>0.2660728</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.686783</td>
<td>2.716203</td>
<td>0.686783</td>
<td>T</td>
<td>0.5211077</td>
<td>0.2892390</td>
<td>T</td>
<td>0.2660728</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.5121312</td>
<td>2.716203</td>
<td>0.2512131</td>
<td>T</td>
<td>0.2921097</td>
<td>0.2805050</td>
<td>T</td>
<td>0.2660728</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.4209181</td>
<td>2.716203</td>
<td>0.4209181</td>
<td>T</td>
<td>0.3480695</td>
<td>0.2973605</td>
<td>T</td>
<td>0.2660728</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.034307</td>
<td>2.716203</td>
<td>0.034307</td>
<td>F</td>
<td>0.1703935</td>
<td>0.3493912</td>
<td>F</td>
<td>0.2660728</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.266072</td>
<td>2.716203</td>
<td>0.266072</td>
<td>F</td>
<td>0.4105325</td>
<td>0.3112040</td>
<td>T</td>
<td>0.2660728</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
Figure 5.9: Den3A-Den15 SS score value results
```r
> source("C:\\Research PHYS Evidence\\personal\\Allie Flores\\Code\\R files\\Mahalanobis outliers\\SO-00 2S.v2.r")
> x=read.table("M1 spectra.csv", sep="", header=T)
> y=read.table("M2 spectra.csv", sep="", header=T)

### Sample 1

<table>
<thead>
<tr>
<th>M1A010T01F1</th>
<th>M1A010T01F2</th>
<th>M1A010T02F1</th>
<th>M1A010T02F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.299512</td>
<td>10.299512</td>
<td>10.299512</td>
<td>10.299512</td>
</tr>
<tr>
<td>10.066564</td>
<td>10.066564</td>
<td>10.066564</td>
<td>10.066564</td>
</tr>
<tr>
<td>11.433212</td>
<td>11.433212</td>
<td>11.433212</td>
<td>11.433212</td>
</tr>
<tr>
<td>8.421313</td>
<td>8.421313</td>
<td>8.421313</td>
<td>8.421313</td>
</tr>
<tr>
<td>10.909724</td>
<td>10.909724</td>
<td>10.909724</td>
<td>10.909724</td>
</tr>
<tr>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
</tr>
<tr>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
</tr>
<tr>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
</tr>
</tbody>
</table>

### Sample 2

<table>
<thead>
<tr>
<th>M1A010T01F1</th>
<th>M1A010T01F2</th>
<th>M1A010T02F1</th>
<th>M1A010T02F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.299512</td>
<td>10.299512</td>
<td>10.299512</td>
<td>10.299512</td>
</tr>
<tr>
<td>10.066564</td>
<td>10.066564</td>
<td>10.066564</td>
<td>10.066564</td>
</tr>
<tr>
<td>11.433212</td>
<td>11.433212</td>
<td>11.433212</td>
<td>11.433212</td>
</tr>
<tr>
<td>8.421313</td>
<td>8.421313</td>
<td>8.421313</td>
<td>8.421313</td>
</tr>
<tr>
<td>10.909724</td>
<td>10.909724</td>
<td>10.909724</td>
<td>10.909724</td>
</tr>
<tr>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
</tr>
<tr>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
</tr>
<tr>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
</tr>
</tbody>
</table>
```

```r
> source("C:\\Research PHYS Evidence\\personal\\Allie Flores\\Code\\R files\\Mahalanobis outliers\\SO-00 2S.v2.r")
> x=read.table("M2 spectra.csv", sep="", header=T)
> y=read.table("M2 spectra.csv", sep="", header=T)

### Sample 1

<table>
<thead>
<tr>
<th>M2A010T01F1</th>
<th>M2A010T01F2</th>
<th>M2A010T02F1</th>
<th>M2A010T02F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.885007</td>
<td>6.885007</td>
<td>6.885007</td>
<td>6.885007</td>
</tr>
<tr>
<td>8.410234</td>
<td>8.410234</td>
<td>8.410234</td>
<td>8.410234</td>
</tr>
<tr>
<td>10.410687</td>
<td>10.410687</td>
<td>10.410687</td>
<td>10.410687</td>
</tr>
<tr>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
</tr>
<tr>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
</tr>
<tr>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
</tr>
</tbody>
</table>

### Sample 2

<table>
<thead>
<tr>
<th>M2A010T01F1</th>
<th>M2A010T01F2</th>
<th>M2A010T02F1</th>
<th>M2A010T02F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.885007</td>
<td>6.885007</td>
<td>6.885007</td>
<td>6.885007</td>
</tr>
<tr>
<td>8.410234</td>
<td>8.410234</td>
<td>8.410234</td>
<td>8.410234</td>
</tr>
<tr>
<td>10.410687</td>
<td>10.410687</td>
<td>10.410687</td>
<td>10.410687</td>
</tr>
<tr>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
<td>10.488472</td>
</tr>
<tr>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
<td>10.778529</td>
</tr>
<tr>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
<td>10.747666</td>
</tr>
</tbody>
</table>
```
<table>
<thead>
<tr>
<th>sampleID</th>
<th>m10_spectra.csv</th>
<th>sep=&quot;&quot;</th>
<th>header=T</th>
</tr>
</thead>
</table>

```r
SD002sample(x, y, qtile=0.95, rm = "0", cent="mc", pctl=95, or = "", alpha=0.05, s500gamma=0) sampleID m10_spectra.csv sep="" header=T
```
Figure 5.10: M1-M6 DS score value results
```r
# source("C:\Research PWS Evidence\Personal\Allie Flores\code\R files\Mahalanobis outliers\50-00 25.2v2.r")

x=read.table("F 1-5 spectra.csv", sep="", header=T)
y=read.table("I 1-5 spectra.csv", sep="", header=T)

# S0002sample(x, y, qtile=0.075, nsim=0, cent="mc", mp=0.95, or="" , alphamc=0.75 , s000dgamma=0)
sampleD Mahalanobis NPC SD Cutoff Score Dist SD disp blanki orthogonal Dist OD Cutoff Odds disp blank Score value S000 disp
1 1 2.629098 2.713260 2.719280 2.629098 0.403919 0.984057 T 0.867983 T
2 1 1.661490 2.713260 1.588654 1.661490 0.601900 0.984057 T 0.584182 T
3 1 3.829564 2.713260 1.970704 3.829564 0.318107 0.984057 T 0.708689 T
4 1 1.661490 2.713260 1.367060 1.661490 0.289518 0.984057 T 0.593187 T
5 1 0.842835 2.713260 1.882835 0.842835 0.484413 0.984057 T 0.674096 T

# S0002sample(x, y, qtile=0.075, nsim=0, cent="mc", mp=0.95, or="" , alphamc=0.75 , s000dgamma=3)
sampleD Mahalanobis NPC SD Cutoff Score Dist SD disp blanki orthogonal Dist OD Cutoff Odds disp blank Score value S000 disp
1 1 2.629098 2.713260 2.719280 2.629098 0.403919 0.984057 T 2.534841 T
2 1 1.661490 2.713260 1.588654 1.661490 0.601900 0.984057 T 3.348742 T
3 1 3.829564 2.713260 1.970704 3.829564 0.318107 0.984057 T 2.174372 T
4 1 1.661490 2.713260 1.367060 1.661490 0.289518 0.984057 T 1.567173 T
5 1 0.842835 2.713260 1.882835 0.842835 0.484413 0.984057 T 2.690829 T

# S0002sample(x, y, qtile=0.075, nsim=0, cent="mc", mp=0.95, or="" , alphamc=0.75 , s000dgamma=1)
sampleD Mahalanobis NPC SD Cutoff Score Dist SD disp blanki orthogonal Dist OD Cutoff Odds disp blank Score value S000 disp
1 1 2.629098 2.713260 2.719280 2.629098 0.403919 0.984057 T 2.534841 T
2 1 1.661490 2.713260 1.588654 1.661490 0.601900 0.984057 T 3.348742 T
3 1 3.829564 2.713260 1.970704 3.829564 0.318107 0.984057 T 2.174372 T
4 1 1.661490 2.713260 1.367060 1.661490 0.289518 0.984057 T 1.567173 T
5 1 0.842835 2.713260 1.882835 0.842835 0.484413 0.984057 T 2.690829 T

# S0002sample(x, y, qtile=0.075, nsim=0, cent="mc", mp=0.95, or="" , alphamc=0.75 , s000dgamma=5)
sampleD Mahalanobis NPC SD Cutoff Score Dist SD disp blanki orthogonal Dist OD Cutoff Odds disp blank Score value S000 disp
1 1 2.218065 2.713260 2.719280 2.218065 0.654877 0.133901 T 0.816055 T
2 1 1.352529 2.713260 1.592259 1.352529 0.241009 0.133901 T 0.583073 T
3 1 2.106529 2.713260 2.106529 2.106529 0.263408 0.133901 T 1.776850 T
4 1 0.840710 2.713260 0.840710 0.840710 0.205482 0.133901 T 0.590502 T
5 1 0.477346 2.713260 0.477346 0.477346 0.303179 0.133901 T 1.610096 T

# S0002sample(x, y, qtile=0.075, nsim=0, cent="mc", mp=0.95, or="" , alphamc=0.75 , s000dgamma=1)
sampleD Mahalanobis NPC SD Cutoff Score Dist SD disp blanki orthogonal Dist OD Cutoff Odds disp blank Score value S000 disp
1 1 2.218065 2.713260 2.719280 2.218065 0.654877 0.133901 T 2.542987 T
2 1 1.352529 2.713260 1.592259 1.352529 0.241009 0.133901 T 1.078743 T
3 1 2.106529 2.713260 2.106529 2.106529 0.263408 0.133901 T 1.551221 T
4 1 0.840710 2.713260 0.840710 0.840710 0.205482 0.133901 T 0.824559 T
5 1 0.477346 2.713260 0.477346 0.477346 0.303179 0.133901 T 1.072082 T

# S0002sample(x, y, qtile=0.075, nsim=0, cent="mc", mp=0.95, or="" , alphamc=0.75 , s000dgamma=7)
sampleD Mahalanobis NPC SD Cutoff Score Dist SD disp blanki orthogonal Dist OD Cutoff Odds disp blank Score value S000 disp
1 1 2.218065 2.713260 2.719280 2.218065 0.654877 0.133901 T 4.269360 T
2 1 1.352529 2.713260 1.592259 1.352529 0.241009 0.133901 T 1.551221 T
3 1 2.106529 2.713260 2.106529 2.106529 0.263408 0.133901 T 1.551221 T
4 1 0.840710 2.713260 0.840710 0.840710 0.205482 0.133901 T 1.551221 T
5 1 0.477346 2.713260 0.477346 0.477346 0.303179 0.133901 T 1.551221 T
```
205
Figure 5.11: F-J DS score value results
```r
# Load data
source("C:\\research\pms\evidence\\personal\\allie-flores\\code\\r files\\mahalanobis outliers \sd-od 2s_v2.r")

# Read tables
xc.read.table("Den3A indiv.csv", sep="/", header=T)
xc.read.table("Den3 indiv.csv", sep="/", header=T)

# Sample data
SD02sample(x, y, qtl=0.975, rm = "0", cent="mc", npcf=.95, or = "", alphec=.05, 000gamma=0)
sampled Mahalanobis NPC SD02cutoff Score Dst SD0 DS blank1 Orthogonal Dst OD Cutoff ODDS Blank Score Value SD0 DS
1 Den3A1F1 1.3094636 2.716203 1.0944036 1.602126 0.6379955 0.463994 F
2 Den3A1F2 0.3068567 2.716203 0.3068567 0.507056 0.6379955 0.410275 F
3 Den3A2F1 1.0360175 2.716203 1.0360175 0.903693 0.6379955 T 0.4806774 F
4 Den3A2F2 0.1234644 2.716203 0.1234644 0.8208556 0.6379955 T 0.3767776 F
5 Den3A3F1 1.7867724 2.716203 1.7867724 0.667479 0.6379955 T 0.581414 F
6 Den3A3F2 1.7867724 2.716203 1.7867724 0.667479 0.6379955 T 0.581414 F
7 Den3A4F1 0.3102759 2.716203 0.3102759 0.885790 0.6379955 T 0.1878687 F
8 Den3A4F2 0.3102759 2.716203 0.3102759 0.885790 0.6379955 T 0.1878687 F

SD02sample(x, y, qtl=0.975, rm = "0", cent="mc", npcf=.95, or = "", alphec=.05, 000gamma=5)
sampled Mahalanobis NPC SD02cutoff Score Dst SD0 DS blank1 Orthogonal Dst OD Cutoff ODDS Blank Score Value SD0 DS
1 Den3A1F1 1.3094636 2.716203 1.0944036 0.602116 0.6379955 0.7125914 F
2 Den3A1F2 0.3068567 2.716203 0.3068567 0.507056 0.6379955 0.7125914 F
3 Den3A2F1 1.0360175 2.716203 1.0360175 0.903693 0.6379955 T 0.4806774 F
4 Den3A2F2 0.1234644 2.716203 0.1234644 0.8208556 0.6379955 T 0.3767776 F
5 Den3A3F1 1.7867724 2.716203 1.7867724 0.667479 0.6379955 T 0.581414 F
6 Den3A3F2 1.7867724 2.716203 1.7867724 0.667479 0.6379955 T 0.581414 F
7 Den3A4F1 0.3102759 2.716203 0.3102759 0.885790 0.6379955 T 0.1878687 F
8 Den3A4F2 0.3102759 2.716203 0.3102759 0.885790 0.6379955 T 0.1878687 F

SD02sample(x, y, qtl=0.975, rm = "0", cent="mc", npcf=.95, or = "", alphec=.05, 000gamma=5)
sampled Mahalanobis NPC SD02cutoff Score Dst SD0 DS blank1 Orthogonal Dst OD Cutoff ODDS Blank Score Value SD0 DS
1 Den3A1F1 1.3094636 2.716203 1.0944036 0.602116 0.6379955 0.7125914 F
2 Den3A1F2 0.3068567 2.716203 0.3068567 0.507056 0.6379955 0.7125914 F
3 Den3A2F1 1.0360175 2.716203 1.0360175 0.903693 0.6379955 T 0.4806774 F
4 Den3A2F2 0.1234644 2.716203 0.1234644 0.8208556 0.6379955 T 0.3767776 F
5 Den3A3F1 1.7867724 2.716203 1.7867724 0.667479 0.6379955 T 0.581414 F
6 Den3A3F2 1.7867724 2.716203 1.7867724 0.667479 0.6379955 T 0.581414 F
7 Den3A4F1 0.3102759 2.716203 0.3102759 0.885790 0.6379955 T 0.1878687 F
8 Den3A4F2 0.3102759 2.716203 0.3102759 0.885790 0.6379955 T 0.1878687 F

# Source
source("C:\\research\pms\evidence\\personal\\allie-flores\\code\\r files\\mahalanobis outliers \sd-od 2s_v2.r")
```
```r
> source("C:\\research PHYS Evidence\\Allie Flores\\code\\R files\\mahalanobis outliers SD-00 25.v2.r")
> x=read.table("den indiv.csv", sep="", header=T)
> y=read.table("dena indiv.csv", sep="", header=T)

```
```r
source("C:\\\Research\PHYS Evidence\\Personal\\Allie Flores\\Code\\R files\\Mahalanobis outliers\\50-00 25v2.r")
```

```r
dx=read.table("bend indiv.csv", sep="", header=T)
```
```r
# Load data
source("C:\\research\\MNS\Evidence\\Personal\\allel_flores\\code\\files\\mahalanobis_outliers_\sd-od_25_v2.r")

# Read tables
require("reshape")
require("dplyr")
require("tidyr")
require("ggplot2")

# Read table "benindiv.csv"
benind <- read.csv("benindiv.csv")

# Read table "benindiv.csv" again
benind <- read.csv("benindiv.csv")

# Sample dataframe
sampled <- benind %>% sample_n(10)

# Print sample dataframe
sampled

# Create a new dataframe
new_df <- sampled %>% mutate(NEW = 1)

# Print new dataframe
new_df

# Load data
source("C:\\research\\MNS\Evidence\\Personal\\allel_flores\\code\\files\\mahalanobis_outliers_\sd-od_25_v2.r")

# Read tables
require("reshape")
require("dplyr")
require("tidyr")
require("ggplot2")

# Read table "benindiv.csv"
benind <- read.csv("benindiv.csv")

# Read table "benindiv.csv" again
benind <- read.csv("benindiv.csv")

# Sample dataframe
sampled <- benind %>% sample_n(10)

# Print sample dataframe
sampled

# Create a new dataframe
new_df <- sampled %>% mutate(NEW = 1)

# Print new dataframe
new_df
```

The text appears to be R code snippets for data manipulation and analysis. It includes commands for loading data, reading tables, sampling, and creating new data frames. The code is designed to work with datasets named `benindiv.csv` and `mahalanobis_outliers_sd-od_25_v2.r`. The output of these operations is not shown in the image, as it is not visualized here. The code seems to be part of a data analysis workflow, possibly for statistical analysis or data visualization.
```r
> source("C:\\Research PHYS Evidence\Personal\Allie Flores\(code\R files\mahalanobis outliers\SD-02 2Svlv.r")
> x=read.table("Derin indiv.csv", sep="", header=T)
> y=read.table("Derin indiv.csv", sep="", header=T)

```
```r
# Import the data
source("C:\Research\PHY Evidence\Personal\Allie Flores\Code\R files\Mahalanobis outliers\SD-OD 25_v2.r")

# Read the data
x <- read.table("denb indiv.csv", sep="", header=T)
y <- read.table("denbA indiv.csv", sep="", header=T)

# Apply the Mahalanobis distance
sdodsample(x = y, qtile = 0.975, mm = "0", cent = "nc", nref = 0.5, or = "", alphamd = 0.75, SDODgamma = 0)

# Sample Mahalanobis NPC SDOD cutoff score dist SD DIs blank 1 orthogonal dist OD cutoff odds blank score value SDOD DIs
# Dep3Aa1F1 1.3611861 2.716203 1.3611861 P 0.4396941 0.709203 F 0.857263 F
# Dep3Aa1F3 1.3312353 2.716203 1.3312353 T 0.5425269 0.709203 T 0.66808 F
# Dep3Aa2F1 1.8403504 2.716203 1.8403504 F 0.5113587 0.709203 T 0.688088 F
# Dep3Aa2F2 0.5249279 2.716203 0.5249279 F 0.6183022 0.709203 T 0.772262 T
# Dep3Aa2F3 1.3312353 2.716203 1.3312353 F 0.8572631 0.709203 T 0.8572631 F
# Dep3Aa2F4 0.4596005 2.716203 0.4596005 F 0.6183022 0.709203 F 0.6183022 F
# Dep3Aa4F1 2.7172407 2.716203 2.7172407 T 1.367915 0.709203 T 1.637794 T
# Dep3Aa4F2 2.7777998 2.716203 2.7777998 T 1.637915 0.709203 T 1.637794 T

# Apply the Mahalanobis distance
sdodsample(x = y, qtile = 0.975, mm = "0", cent = "nc", nref = 0.5, or = "", alphamd = 0.75, SDODgamma = 1)

# Sample Mahalanobis NPC SDOD cutoff score dist SD DIs blank 1 orthogonal dist OD cutoff odds blank score value SDOD DIs
# Dep3Aa1F1 1.3611861 2.716203 1.3611861 F 0.4396941 0.709203 T 0.6119096 F
# Dep3Aa1F3 1.3312353 2.716203 1.3312353 T 0.5425269 0.709203 T 0.857263 F
# Dep3Aa2F1 1.8403504 2.716203 1.8403504 F 0.5113587 0.709203 T 0.688088 F
# Dep3Aa2F2 0.5249279 2.716203 0.5249279 F 0.6183022 0.709203 T 0.772262 T
# Dep3Aa2F3 1.3312353 2.716203 1.3312353 F 0.8572631 0.709203 T 0.8572631 F
# Dep3Aa2F4 0.4596005 2.716203 0.4596005 F 0.6183022 0.709203 F 0.6183022 F
# Dep3Aa4F1 2.7172407 2.716203 2.7172407 T 1.367915 0.709203 T 1.637794 T
# Dep3Aa4F2 2.7777998 2.716203 2.7777998 T 1.637915 0.709203 T 1.637794 T

# Apply the Mahalanobis distance
sdodsample(x = y, qtile = 0.975, mm = "0", cent = "nc", nref = 0.5, or = "", alphamd = 0.75, SDODgamma = 0)

# Sample Mahalanobis NPC SDOD cutoff score dist SD DIs blank 1 orthogonal dist OD cutoff odds blank score value SDOD DIs
# Dep3Aa1F1 1.3611861 2.716203 1.3611861 F 0.4396941 0.709203 T 0.857263 F
# Dep3Aa1F3 1.3312353 2.716203 1.3312353 T 0.5425269 0.709203 T 0.8572631 F
# Dep3Aa2F1 1.8403504 2.716203 1.8403504 F 0.5113587 0.709203 T 0.688088 F
# Dep3Aa2F2 0.5249279 2.716203 0.5249279 F 0.6183022 0.709203 T 0.772262 T
# Dep3Aa2F3 1.3312353 2.716203 1.3312353 T 0.8572631 0.709203 F 0.8572631 F
# Dep3Aa2F4 0.4596005 2.716203 0.4596005 F 0.6183022 0.709203 F 0.6183022 F
# Dep3Aa4F1 2.7172407 2.716203 2.7172407 T 1.367915 0.709203 T 1.637794 T
# Dep3Aa4F2 2.7777998 2.716203 2.7777998 T 1.637915 0.709203 T 1.637794 T

# Apply the Mahalanobis distance
sdodsample(x = y, qtile = 0.975, mm = "0", cent = "nc", nref = 0.5, or = "", alphamd = 0.75, SDODgamma = 1)

# Sample Mahalanobis NPC SDOD cutoff score dist SD DIs blank 1 orthogonal dist OD cutoff odds blank score value SDOD DIs
# Dep3Aa1F1 1.3611861 2.716203 1.3611861 F 0.4396941 0.709203 T 0.6119096 F
# Dep3Aa1F3 1.3312353 2.716203 1.3312353 T 0.5425269 0.709203 T 0.857263 F
# Dep3Aa2F1 1.8403504 2.716203 1.8403504 F 0.5113587 0.709203 T 0.688088 F
# Dep3Aa2F2 0.5249279 2.716203 0.5249279 F 0.6183022 0.709203 T 0.772262 T
# Dep3Aa2F3 1.3312353 2.716203 1.3312353 F 0.8572631 0.709203 T 0.8572631 F
# Dep3Aa2F4 0.4596005 2.716203 0.4596005 F 0.6183022 0.709203 F 0.6183022 F
# Dep3Aa4F1 2.7172407 2.716203 2.7172407 T 1.367915 0.709203 T 1.637794 T
# Dep3Aa4F2 2.7777998 2.716203 2.7777998 T 1.637915 0.709203 T 1.637794 T

# Source
source("C:\Research\PHY Evidence\Personal\Allie Flores\Code\R files\Mahalanobis outliers\SD-OD 25_v2.r")
```

216
<table>
<thead>
<tr>
<th>Sampled MalahanobisNPC SD CutOff Score Dist SD Dis blank</th>
<th>Orthogonal Dist OD CutOff Odds Dist blank</th>
<th>Score Value SDOD Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Den5A1F2</td>
<td>2.167294</td>
<td>2.167294</td>
</tr>
<tr>
<td>2 Den5A2F2</td>
<td>1.329860</td>
<td>1.329860</td>
</tr>
<tr>
<td>3 Den5A3F2</td>
<td>1.021229</td>
<td>1.021229</td>
</tr>
<tr>
<td>4 Den5A4F2</td>
<td>2.539535</td>
<td>2.539535</td>
</tr>
<tr>
<td>5 Den5A5F2</td>
<td>2.012458</td>
<td>2.012458</td>
</tr>
<tr>
<td>6 Den5A6F2</td>
<td>1.539300</td>
<td>1.539300</td>
</tr>
<tr>
<td>7 Den5A7F2</td>
<td>1.398600</td>
<td>1.398600</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampled MalahanobisNPC SD CutOff Score Dist SD Dis blank</th>
<th>Orthogonal Dist OD CutOff Odds Dist blank</th>
<th>Score Value SDOD Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Den4A1F1</td>
<td>2.220050</td>
<td>2.220050</td>
</tr>
<tr>
<td>2 Den4A2F1</td>
<td>1.617524</td>
<td>1.617524</td>
</tr>
<tr>
<td>3 Den4A3F1</td>
<td>1.329860</td>
<td>1.329860</td>
</tr>
<tr>
<td>4 Den4A4F1</td>
<td>1.021229</td>
<td>1.021229</td>
</tr>
<tr>
<td>5 Den4A5F1</td>
<td>2.539535</td>
<td>2.539535</td>
</tr>
<tr>
<td>6 Den4A6F1</td>
<td>2.012458</td>
<td>2.012458</td>
</tr>
<tr>
<td>7 Den4A7F1</td>
<td>1.539300</td>
<td>1.539300</td>
</tr>
<tr>
<td>8 Den4A8F1</td>
<td>1.398600</td>
<td>1.398600</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampled MalahanobisNPC SD CutOff Score Dist SD Dis blank</th>
<th>Orthogonal Dist OD CutOff Odds Dist blank</th>
<th>Score Value SDOD Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Den3A1F1</td>
<td>2.220050</td>
<td>2.220050</td>
</tr>
<tr>
<td>2 Den3A2F1</td>
<td>1.617524</td>
<td>1.617524</td>
</tr>
<tr>
<td>3 Den3A3F1</td>
<td>1.329860</td>
<td>1.329860</td>
</tr>
<tr>
<td>4 Den3A4F1</td>
<td>1.021229</td>
<td>1.021229</td>
</tr>
<tr>
<td>5 Den3A5F1</td>
<td>2.539535</td>
<td>2.539535</td>
</tr>
<tr>
<td>6 Den3A6F1</td>
<td>2.012458</td>
<td>2.012458</td>
</tr>
<tr>
<td>7 Den3A7F1</td>
<td>1.539300</td>
<td>1.539300</td>
</tr>
<tr>
<td>8 Den3A8F1</td>
<td>1.398600</td>
<td>1.398600</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampled MalahanobisNPC SD CutOff Score Dist SD Dis blank</th>
<th>Orthogonal Dist OD CutOff Odds Dist blank</th>
<th>Score Value SDOD Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Den2A1F1</td>
<td>2.220050</td>
<td>2.220050</td>
</tr>
<tr>
<td>2 Den2A2F1</td>
<td>1.617524</td>
<td>1.617524</td>
</tr>
<tr>
<td>3 Den2A3F1</td>
<td>1.329860</td>
<td>1.329860</td>
</tr>
<tr>
<td>4 Den2A4F1</td>
<td>1.021229</td>
<td>1.021229</td>
</tr>
<tr>
<td>5 Den2A5F1</td>
<td>2.539535</td>
<td>2.539535</td>
</tr>
<tr>
<td>6 Den2A6F1</td>
<td>2.012458</td>
<td>2.012458</td>
</tr>
<tr>
<td>7 Den2A7F1</td>
<td>1.539300</td>
<td>1.539300</td>
</tr>
<tr>
<td>8 Den2A8F1</td>
<td>1.398600</td>
<td>1.398600</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampled MalahanobisNPC SD CutOff Score Dist SD Dis blank</th>
<th>Orthogonal Dist OD CutOff Odds Dist blank</th>
<th>Score Value SDOD Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Den1A1F1</td>
<td>2.220050</td>
<td>2.220050</td>
</tr>
<tr>
<td>2 Den1A2F1</td>
<td>1.617524</td>
<td>1.617524</td>
</tr>
<tr>
<td>3 Den1A3F1</td>
<td>1.329860</td>
<td>1.329860</td>
</tr>
<tr>
<td>4 Den1A4F1</td>
<td>1.021229</td>
<td>1.021229</td>
</tr>
<tr>
<td>5 Den1A5F1</td>
<td>2.539535</td>
<td>2.539535</td>
</tr>
<tr>
<td>6 Den1A6F1</td>
<td>2.012458</td>
<td>2.012458</td>
</tr>
<tr>
<td>7 Den1A7F1</td>
<td>1.539300</td>
<td>1.539300</td>
</tr>
<tr>
<td>8 Den1A8F1</td>
<td>1.398600</td>
<td>1.398600</td>
</tr>
</tbody>
</table>
```r
> source("/\Research\PHYS\Evidence\\Personal\Allie\Flores\\Code\\R\files\\mahalanobis\outliers\\50-00\2.\v2.r")
> x=read.table("denb indiv.csv", sep=";", header=T)
> y=read.table("den3A indiv.csv", sep=";", header=T)

```
<table>
<thead>
<tr>
<th>SampleID</th>
<th>MahalanobisDist</th>
<th>NPC_Score</th>
<th>Score</th>
<th>Dist</th>
<th>SD_Score</th>
<th>Blank1</th>
<th>Orthogonal</th>
<th>Dist_Cutoff</th>
<th>Odds</th>
<th>Blank</th>
<th>Score_Value</th>
<th>SD00_Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.4128915</td>
<td>2.716203</td>
<td>3.4386313</td>
<td>T</td>
<td>1.809167</td>
<td>0.254145</td>
<td>T</td>
<td>1.2359641</td>
<td>T</td>
<td>0.3926098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.3397160</td>
<td>2.716203</td>
<td>4.3591600</td>
<td>T</td>
<td>2.829001</td>
<td>0.254145</td>
<td>T</td>
<td>1.6712772</td>
<td>T</td>
<td>0.4311818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3079314</td>
<td>2.716203</td>
<td>0.3079314</td>
<td>F</td>
<td>1.375212</td>
<td>0.254145</td>
<td>T</td>
<td>0.6275187</td>
<td>F</td>
<td>0.0000006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.7776552</td>
<td>2.716203</td>
<td>2.7776552</td>
<td>T</td>
<td>1.673323</td>
<td>0.254145</td>
<td>T</td>
<td>1.0226790</td>
<td>T</td>
<td>0.4255293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.8735227</td>
<td>2.716203</td>
<td>3.8735227</td>
<td>T</td>
<td>2.987567</td>
<td>0.254145</td>
<td>T</td>
<td>1.4268080</td>
<td>T</td>
<td>0.5415566</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.0264448</td>
<td>2.716203</td>
<td>3.0264448</td>
<td>T</td>
<td>2.467351</td>
<td>0.254145</td>
<td>T</td>
<td>0.8453401</td>
<td>T</td>
<td>0.2457702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.2204632</td>
<td>2.716203</td>
<td>4.2204632</td>
<td>T</td>
<td>2.356236</td>
<td>0.254145</td>
<td>T</td>
<td>1.3539099</td>
<td>T</td>
<td>0.2480089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4.4900754</td>
<td>2.716203</td>
<td>4.4900754</td>
<td>T</td>
<td>2.262097</td>
<td>0.254145</td>
<td>T</td>
<td>1.6538790</td>
<td>T</td>
<td>0.3223156</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SampleID</th>
<th>MahalanobisDist</th>
<th>NPC_Score</th>
<th>Score</th>
<th>Dist</th>
<th>SD_Score</th>
<th>Blank1</th>
<th>Orthogonal</th>
<th>Dist_Cutoff</th>
<th>Odds</th>
<th>Blank</th>
<th>Score_Value</th>
<th>SD00_Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.480155</td>
<td>2.716203</td>
<td>4.480155</td>
<td>T</td>
<td>1.809167</td>
<td>0.254145</td>
<td>T</td>
<td>4.102998</td>
<td>T</td>
<td>0.4311818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.3397160</td>
<td>2.716203</td>
<td>4.3397160</td>
<td>T</td>
<td>2.829001</td>
<td>0.254145</td>
<td>T</td>
<td>1.6712772</td>
<td>T</td>
<td>0.4311818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3079314</td>
<td>2.716203</td>
<td>0.3079314</td>
<td>F</td>
<td>1.375212</td>
<td>0.254145</td>
<td>T</td>
<td>0.6275187</td>
<td>F</td>
<td>0.0000006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.7776552</td>
<td>2.716203</td>
<td>2.7776552</td>
<td>T</td>
<td>1.673323</td>
<td>0.254145</td>
<td>T</td>
<td>1.0226790</td>
<td>T</td>
<td>0.4255293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.8735227</td>
<td>2.716203</td>
<td>3.8735227</td>
<td>T</td>
<td>2.987567</td>
<td>0.254145</td>
<td>T</td>
<td>1.4268080</td>
<td>T</td>
<td>0.5415566</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.0264448</td>
<td>2.716203</td>
<td>3.0264448</td>
<td>T</td>
<td>2.467351</td>
<td>0.254145</td>
<td>T</td>
<td>0.8453401</td>
<td>T</td>
<td>0.2457702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.2204632</td>
<td>2.716203</td>
<td>4.2204632</td>
<td>T</td>
<td>2.356236</td>
<td>0.254145</td>
<td>T</td>
<td>1.3539099</td>
<td>T</td>
<td>0.2480089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4.4900754</td>
<td>2.716203</td>
<td>4.4900754</td>
<td>T</td>
<td>2.262097</td>
<td>0.254145</td>
<td>T</td>
<td>1.6538790</td>
<td>T</td>
<td>0.3223156</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SampleID</th>
<th>MahalanobisDist</th>
<th>NPC_Score</th>
<th>Score</th>
<th>Dist</th>
<th>SD_Score</th>
<th>Blank1</th>
<th>Orthogonal</th>
<th>Dist_Cutoff</th>
<th>Odds</th>
<th>Blank</th>
<th>Score_Value</th>
<th>SD00_Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.4128915</td>
<td>2.716203</td>
<td>3.4128915</td>
<td>T</td>
<td>1.809167</td>
<td>0.254145</td>
<td>T</td>
<td>7.118021</td>
<td>T</td>
<td>0.3926098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.3397160</td>
<td>2.716203</td>
<td>4.3397160</td>
<td>T</td>
<td>2.829001</td>
<td>0.254145</td>
<td>T</td>
<td>11.107804</td>
<td>T</td>
<td>0.4311818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3079314</td>
<td>2.716203</td>
<td>0.3079314</td>
<td>F</td>
<td>1.375212</td>
<td>0.254145</td>
<td>T</td>
<td>5.431118</td>
<td>F</td>
<td>0.0000006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.7776552</td>
<td>2.716203</td>
<td>2.7776552</td>
<td>T</td>
<td>1.673323</td>
<td>0.254145</td>
<td>T</td>
<td>6.587681</td>
<td>T</td>
<td>0.4255293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.8735227</td>
<td>2.716203</td>
<td>3.8735227</td>
<td>T</td>
<td>2.987567</td>
<td>0.254145</td>
<td>T</td>
<td>8.745181</td>
<td>T</td>
<td>0.5415566</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.0264448</td>
<td>2.716203</td>
<td>3.0264448</td>
<td>T</td>
<td>2.467351</td>
<td>0.254145</td>
<td>T</td>
<td>8.526537</td>
<td>T</td>
<td>0.2457702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.2204632</td>
<td>2.716203</td>
<td>4.2204632</td>
<td>T</td>
<td>2.356236</td>
<td>0.254145</td>
<td>T</td>
<td>9.182225</td>
<td>T</td>
<td>0.2480089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4.4900754</td>
<td>2.716203</td>
<td>4.4900754</td>
<td>T</td>
<td>2.262097</td>
<td>0.254145</td>
<td>T</td>
<td>8.300789</td>
<td>T</td>
<td>0.3223156</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```R
# Load necessary library
library(MASS)

# Read data
data <- read.csv('denh_indiv.csv', sep=';', header=T)

# Perform Mahalanobis distance
mahalanobis_outliers <- function(data, alpha=0.05) {
  # Calculate Mahalanobis distances
  distances <- Mahalanobis(data, center=colMeans(data), cov=vcov(data))

  # Get cutoff value
  cutoff <- qchisq(1-alpha, df=ncol(data)-1)

  # Identify outliers
  outliers <- which(distances > cutoff)

  # Return outliers
  return(outliers)
}

# Apply function
outliers <- mahalanobis_outliers(data, alpha=0.05)

# Print outliers
outliers
```

**Output:**

```
[1]   2   3   4   5   6   7   8
```

221
```r
> source("C:\\Research\PHS_Evidence\Personal\\Allie Flores\\code\\files\\mahalanobis outliers\\SD-00 2S_2v2.r")
> x=read.table("Denn indiv.csv", sep="", header=T)
> y=read.table("Denn A indiv.csv", sep="", header=T)

<table>
<thead>
<tr>
<th>SampleID</th>
<th>Mahalanobis Dist</th>
<th>NPC</th>
<th>ECutoff</th>
<th>Score</th>
<th>Dist</th>
<th>SD</th>
<th>Dis</th>
<th>Blank</th>
<th>Orthogonal</th>
<th>Dist</th>
<th>SD</th>
<th>Cutoff</th>
<th>Odds</th>
<th>Blank</th>
<th>Score</th>
<th>Value</th>
<th>SD</th>
<th>Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
<table>
<thead>
<tr>
<th>SampleID</th>
<th>Mahalanobis npc SC cutoff</th>
<th>Score dist SD dist</th>
<th>blank1 orthogonal1 dist OD cutoff</th>
<th>OD dist blank1 orth OD</th>
<th>Score value</th>
<th>SD OD dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>S00D1201</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1202</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1203</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1204</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1205</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1206</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1207</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
<tr>
<td>S00D1208</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
<td>0.223</td>
<td>1.25571</td>
<td>1.25571</td>
</tr>
</tbody>
</table>
null
```r
# Load the data
x <- read.table("Denn5 indiv.csv", sep="", header=T)
y <- read.table("Denn3 indiv.csv", sep="", header=T)

# Sample data
x_sample <- x[x$y < 0.975,]

# Sample data with cut-off
x_sample_1 <- x_sample[x_sample$p > 0.95,]

# Sample data with all values
x_sample_2 <- x_sample

# Print the sample data
print(x_sample_1)
print(x_sample_2)
```

```
   y   x   p   q
1 0.2 0.9 0.95 0.8
2 0.3 0.8 0.90 0.7
3 0.4 0.7 0.85 0.6
4 0.5 0.6 0.80 0.5
5 0.6 0.5 0.75 0.4
6 0.7 0.4 0.70 0.3
7 0.8 0.3 0.65 0.2
8 0.9 0.2 0.60 0.1

# Print the sample data with cut-off
print(x_sample_2)

   y   x   p   q
1 0.2 0.9 0.95 0.8
2 0.3 0.8 0.90 0.7
3 0.4 0.7 0.85 0.6
4 0.5 0.6 0.80 0.5
5 0.6 0.5 0.75 0.4
6 0.7 0.4 0.70 0.3
7 0.8 0.3 0.65 0.2
8 0.9 0.2 0.60 0.1
```

```
# Load the data
x <- read.table("Denn5 indiv.csv", sep="", header=T)
y <- read.table("Denn3 indiv.csv", sep="", header=T)

# Sample data
x_sample <- x[x$y < 0.975,]

# Sample data with cut-off
x_sample_1 <- x_sample[x_sample$p > 0.95,]

# Sample data with all values
x_sample_2 <- x_sample

# Print the sample data
print(x_sample_1)
print(x_sample_2)
```

```
   y   x   p   q
1 0.2 0.9 0.95 0.8
2 0.3 0.8 0.90 0.7
3 0.4 0.7 0.85 0.6
4 0.5 0.6 0.80 0.5
5 0.6 0.5 0.75 0.4
6 0.7 0.4 0.70 0.3
7 0.8 0.3 0.65 0.2
8 0.9 0.2 0.60 0.1

# Print the sample data with cut-off
print(x_sample_2)

   y   x   p   q
1 0.2 0.9 0.95 0.8
2 0.3 0.8 0.90 0.7
3 0.4 0.7 0.85 0.6
4 0.5 0.6 0.80 0.5
5 0.6 0.5 0.75 0.4
6 0.7 0.4 0.70 0.3
7 0.8 0.3 0.65 0.2
8 0.9 0.2 0.60 0.1
```
```r
# Source: "C:\\Research PHYS Evidence\\alle flores\\Code\\R files\\Mahalanobis outliers\\so-d0 DS_v2.r"
# x-read.table("ben5 indiv.csv", sep = "", header = T)
# y-read.table("ben5 indiv.csv", sep = "", header = T)

SD0Dsample(x, y, qtile = 0.975, mm = "0", cent = "mc", rpcr = 0.95, OR = "", alphahcd = 0.75, SSD0gamma = 0)

<table>
<thead>
<tr>
<th>sampleID</th>
<th>Mahalanobis</th>
<th>NPC</th>
<th>SSD0ctt</th>
<th>Score</th>
<th>Dist</th>
<th>SSD0</th>
<th>Dist</th>
<th>Blank</th>
<th>orthogonal</th>
<th>Dist</th>
<th>OOD</th>
<th>cutoft</th>
<th>OOD</th>
<th>OOD</th>
<th>cutoft</th>
<th>Blank</th>
<th>Score</th>
<th>Value</th>
<th>SSD0</th>
<th>Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.9026827</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.678912</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.932597</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.716818</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.002040</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.585097</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.590684</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.110427</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SD0Dsample(x, y, qtile = 0.975, mm = "0", cent = "mc", rpcr = 0.95, OR = "", alphahcd = 0.75, SSD0gamma = 0)

<table>
<thead>
<tr>
<th>sampleID</th>
<th>Mahalanobis</th>
<th>NPC</th>
<th>SSD0ctt</th>
<th>Score</th>
<th>Dist</th>
<th>SSD0</th>
<th>Dist</th>
<th>Blank</th>
<th>orthogonal</th>
<th>Dist</th>
<th>OOD</th>
<th>cutoft</th>
<th>OOD</th>
<th>OOD</th>
<th>cutoft</th>
<th>Blank</th>
<th>Score</th>
<th>Value</th>
<th>SSD0</th>
<th>Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.9026827</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.678912</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.932597</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.716818</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.002040</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.585097</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.590684</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.110427</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: "C:\\Research PHYS Evidence\\alle flores\\Code\\R files\\Mahalanobis outliers\\so-d0 DS_v2.r"

x-read.table("ben5 indiv.csv", sep = "", header = T)
y-read.table("ben5 indiv.csv", sep = "", header = T)

SD0Dsample(x, y, qtile = 0.975, mm = "0", cent = "mc", rpcr = 0.95, OR = "", alphahcd = 0.75, SSD0gamma = 0)

<table>
<thead>
<tr>
<th>sampleID</th>
<th>Mahalanobis</th>
<th>NPC</th>
<th>SSD0ctt</th>
<th>Score</th>
<th>Dist</th>
<th>SSD0</th>
<th>Dist</th>
<th>Blank</th>
<th>orthogonal</th>
<th>Dist</th>
<th>OOD</th>
<th>cutoft</th>
<th>OOD</th>
<th>OOD</th>
<th>cutoft</th>
<th>Blank</th>
<th>Score</th>
<th>Value</th>
<th>SSD0</th>
<th>Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.9026827</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.678912</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.932597</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.716818</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.002040</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.585097</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.590684</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.110427</td>
<td>2.716203</td>
<td>3.067203</td>
<td>1.706944</td>
<td>0.341054</td>
<td>T</td>
<td>2.909363</td>
<td>T</td>
<td>2.916084</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

226
<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Mahalanobis D5 NPC</th>
<th>SDcutoff</th>
<th>Score Dist</th>
<th>SD_Dis</th>
<th>Blank</th>
<th>Orthogonal Dist</th>
<th>OD_Cutoff</th>
<th>Oddis</th>
<th>Blank</th>
<th>Score Value</th>
<th>SDOD_Dis</th>
</tr>
</thead>
<tbody>
<tr>
<td>DenA1F1</td>
<td>4.840765</td>
<td>4.840765</td>
<td>2.716203</td>
<td>3.457754</td>
<td>1.969716</td>
<td>0.7090353</td>
<td>0.410354</td>
<td>T</td>
<td>T</td>
<td>1.320758</td>
<td>0.5410354</td>
</tr>
<tr>
<td>DenA1F2</td>
<td>8.015853</td>
<td>8.015853</td>
<td>2.716203</td>
<td>3.457754</td>
<td>1.969716</td>
<td>0.7090353</td>
<td>0.410354</td>
<td>T</td>
<td>T</td>
<td>1.320758</td>
<td>0.5410354</td>
</tr>
<tr>
<td>DenA2F1</td>
<td>4.145277</td>
<td>4.145277</td>
<td>2.716203</td>
<td>3.457754</td>
<td>1.969716</td>
<td>0.7090353</td>
<td>0.410354</td>
<td>T</td>
<td>T</td>
<td>1.320758</td>
<td>0.5410354</td>
</tr>
<tr>
<td>DenA2F2</td>
<td>4.145277</td>
<td>4.145277</td>
<td>2.716203</td>
<td>3.457754</td>
<td>1.969716</td>
<td>0.7090353</td>
<td>0.410354</td>
<td>T</td>
<td>T</td>
<td>1.320758</td>
<td>0.5410354</td>
</tr>
<tr>
<td>DenA3F1</td>
<td>5.121519</td>
<td>5.121519</td>
<td>2.716203</td>
<td>3.457754</td>
<td>1.969716</td>
<td>0.7090353</td>
<td>0.410354</td>
<td>T</td>
<td>T</td>
<td>1.320758</td>
<td>0.5410354</td>
</tr>
<tr>
<td>DenA3F2</td>
<td>3.008001</td>
<td>3.008001</td>
<td>2.716203</td>
<td>3.457754</td>
<td>1.969716</td>
<td>0.7090353</td>
<td>0.410354</td>
<td>T</td>
<td>T</td>
<td>1.320758</td>
<td>0.5410354</td>
</tr>
</tbody>
</table>

Figure 5.12: Den3A-Den15 DS score value results


73. Oien CT. Forensic Hair Comparison: Background Information for Interpretation. Forensic Science Communications. 2009;11(12).


80. Liszewski EA. Instrumental and Statistical Methods for the Comparison of Class Evidence. Indianapolis, IN: Purdue University, 2010.

81. Szkudlarek CA. Multivariate Statistical Methods Applied to the Analysis of Trace EVidence. Indianapolis, IN: Purdue University, 2012.

82. Reichard EJ. Chemometrics Applied to the Discrimination of Synthetic Fibers by Microspectrophotometry. Indianapolis, IN: Purdue University, 2013.


