Exploring the Impact of Challenging Behaviors on Treatment Efficacy in Autism Spectrum Disorder

Juliana Hoag
Chapman University, julianahoag@gmail.com

Follow this and additional works at: https://digitalcommons.chapman.edu/cads_theses
Part of the Other Psychiatry and Psychology Commons

Recommended Citation
https://digitalcommons.chapman.edu/cads_theses/4

This Thesis is brought to you for free and open access by the Dissertations and Theses at Chapman University Digital Commons. It has been accepted for inclusion in Computational and Data Sciences (MS) Theses by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.
Exploring the Impact of Challenging Behaviors on Treatment Efficacy in Autism Spectrum Disorder

A Thesis by

Juliana Gardner-Hoag

Chapman University,
Orange, CA
Schmid College of Science and Technology
Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Computational and Data Sciences
May 2019

Thesis Committee:
Erik J. Linstead, Ph.D., Co-Chair
Elizabeth Stevens, Ph.D., Co-Chair
Dennis Dixon, Ph.D.
The thesis of Juliana Gardner-Hoag is approved.

Erik Linstead, Ph.D., Co-Chair

Elizabeth Stevens, Ph.D., Co-Chair

Dennis Dixon, Ph.D.

May 2019
Exploring the Impact of Challenging Behaviors on Treatment Efficacy in Autism Spectrum Disorder

Copyright © 2019

by Juliana Gardner-Hoag
ACKNOWLEDGMENTS

I would like to thank my family and friends for supporting me and encouraging me throughout my graduate school experience and my committee members, Dr. Erik Linstead and Dr. Elizabeth Stevens for helping me find a research area that I was interested in. Thank you CARD and Dennis Dixon for providing the data and giving me insight on how this study could positively impact those diagnosed with Autism Spectrum Disorder. And finally, I would like to thank Adrienne Bergh and Chelsea Parlett-Pelleriti along with the rest of MLAT for being supportive friends and colleagues.
VITA

Juliana Gardner-Hoag

EDUCATION

Master of Science in Computational and Data Sciences 2019
Chapman University  Orange, California

Bachelor of Science in Mathematics 2014
University of Redlands  Redlands, California

RESEARCH EXPERIENCE

Graduate Research Assistant 2018–2019
Chapman University  Orange, California
ABSTRACT

Exploring the Impact of Challenging Behaviors on Treatment Efficacy in Autism Spectrum Disorder

by Juliana Gardner-Hoag

The focus of this study was to explore the impact of challenging behaviors on Applied Behaviors Analysis treatment in Autism Spectrum Disorder. The prevalence of ASD is on the rise, so it is important that we understand how patients are responding to treatment. In this study, we cluster patients (N=854) based on their eight observed challenging behaviors using k-means, a machine learning algorithm, and then perform a multiple linear regression analysis to find significant differences between average exemplars mastered. The goal of this study was to expand the research in the area of ABA treatment for ASD and to help provide more insight helpful for creating personalized therapeutic interventions with maximum efficacy, minimum time and minimum cost for individuals.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>VITA</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Autism Spectrum Disorder and Applied Behavior Analysis</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Thesis Guideline References</td>
<td>7</td>
</tr>
<tr>
<td><strong>2 Related Work</strong></td>
<td>8</td>
</tr>
<tr>
<td>2.1 Treatment Intensity and Duration</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Hierarchical Clustering on Patients Diagnosed with ASD</td>
<td>9</td>
</tr>
<tr>
<td><strong>3 Data</strong></td>
<td>11</td>
</tr>
<tr>
<td>3.1 The Center for Autism and Related Disorders</td>
<td>11</td>
</tr>
<tr>
<td>3.2 Database</td>
<td>12</td>
</tr>
</tbody>
</table>
4 Methods

4.1 K-Means Clustering ........................................... 14
4.2 Linear Regression ........................................... 17

5 Results

5.1 Clustering Results ........................................... 20
5.2 Linear Regression Results ................................... 25
5.3 Tukey Post-Hoc .............................................. 29

6 Discussion ..................................................... 34

7 Future Work ..................................................... 36

8 Conclusion ....................................................... 37

References ......................................................... 39

Appendices ......................................................... 41
LIST OF TABLES

Page

2.1 Eight skills used by Linstead et al. in their multiple regression analysis 8

3.1 Challenging Behaviors . . . . . . . . . . . . . . . . . . . . . . . . . . 12

5.1 This table shows the breakdown of how many individuals are in each
cluster and the challenging behavior of the phenotype. . . . . . . . . 22

5.2 Mean Squared Error Comparison . . . . . . . . . . . . . . . . . . . . 26

5.3 Averages for each cluster . . . . . . . . . . . . . . . . . . . . . . . . . 28

5.4 Interactions and P-values . . . . . . . . . . . . . . . . . . . . . . . . . 29

5.5 Averages of therapy hours and exemplars mastered for the females of
each cluster . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

5.6 Averages of therapy hours and exemplars mastered for the males of
each cluster . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Sums of Squared Differences for Females</td>
<td>20</td>
</tr>
<tr>
<td>5.2</td>
<td>Sums of Squared Differences for Males</td>
<td>21</td>
</tr>
<tr>
<td>5.3</td>
<td>Silhouette Plots: Females and Males</td>
<td>21</td>
</tr>
<tr>
<td>5.4</td>
<td>Silhouette Plot for All Patients</td>
<td>22</td>
</tr>
<tr>
<td>5.5</td>
<td>Radar Chart of All Clusters</td>
<td>23</td>
</tr>
<tr>
<td>5.6</td>
<td>Radar Charts for Phenotypes</td>
<td>24</td>
</tr>
<tr>
<td>5.7</td>
<td>Cluster 4 and Cluster 7</td>
<td>25</td>
</tr>
<tr>
<td>5.8</td>
<td>Line-of-Best Fit for Each Cluster</td>
<td>26</td>
</tr>
<tr>
<td>5.9</td>
<td>Female vs Male Regression Lines</td>
<td>27</td>
</tr>
<tr>
<td>5.10</td>
<td>Female vs Male Regression Lines for Each Cluster</td>
<td>28</td>
</tr>
<tr>
<td>5.11</td>
<td>Boxplots of Clusters</td>
<td>29</td>
</tr>
<tr>
<td>5.12</td>
<td>Boxplots of Each Cluster and Gender</td>
<td>30</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Autism Spectrum Disorder and Applied Behavior Analysis

Autism spectrum disorder, commonly referred to as ASD, is a developmental disability diagnosable by observed deficits in social communication and social interaction, and the presence of restricted, repetitive patterns of behavior, interests, or activities that can persist throughout life [1]. Often, there is nothing in appearance that sets those afflicted with ASD apart from those not afflicted. Individuals with ASD may differ from others in the way they communicate, interact, behave and learn. People with ASD can range from gifted to severely challenged in the way that they learn, think, and solve problems.

In April 2018, the Centers for Disease Control and Prevention (CDC) announced that there was a 15% increase in prevalence of ASD in the United States. The prevalence rose to approximately 1 in 59 children from 1 in 68 children, which was the estimated
prevalence two years prior. Additionally, the CDC reported that the gender gap in autism has decreased, meaning that more females are being diagnosed (1 in 37 males versus 1 in 151 females). The ethnic gap has decreased as well [2].

In 1911, German psychiatrist Eugen Bleuler first used the term autism to describe a symptom he noticed in the most severe cases of schizophrenia [9]. He believed that autistic thinking was when an individual would avoid realities and replace them with fantasies. This was an act he called infantile wishing. Bleuler’s definition of autism was about the individual’s inner life and was not obvious to others by appearance.

Autistic disorder, otherwise known as childhood autism or infantile autism, was first explained by Leo Kanner in 1943. His report consisted of his findings from 11 children who appeared to express two features of autism [11]. These features were a lack of interest in the social world and behaviors he called resistance to change.

By 1970 it was known that autism was characterized by impaired language and communication skills, resistance to change (also referred to as insistence on sameness), motor mannerisms and stereotypies. Stereotypies refer to repetitive body movements and vocalizations (repeating phrases out of context, use of jargon).

The Diagnostic and Statistical Manual (DSM) is the handbook that health care professionals in the United States and around the world follow to diagnose mental disorders [6]. In the DSM, there are descriptions and symptoms along with criteria for diagnosing mental disorders. The DSM provides all health care professionals the ability to communicate about patient diagnoses and establish consistent and dependable diagnoses. Having consistent and reliable diagnoses is important because it allows us to do research and build on the research of others with consistent results, which in turn provides validity to analyses. It also allows for research results to influence
future revisions of the DSM and help in the development of new medications and treatments for disorders.

The DSM was first published in 1952, and since its first version there have been many advances made in what we know about mental disorders. Autism was first officially recognized by the Diagnostic Statistical Manual in 1987 when the DSM-III was released [19]. Autism was included in the DSM-III under a class of conditions called pervasive developmental disorder (PDD). This definition of autism focused on infants with autism. Categories for late-onset autism were included in the DSM-III as well, but there was not much explanation.

In 1994, the DSM-IV was released and since its release, research and publications on autism have grown at a phenomenal rate [19]. Before releasing the DSM-IV, there were many preliminary steps taken, including many literature reviews and reanalysis of data. The American Psychiatric Association (APA) wanted the DSM-IV to balance sensitivity and specificity across the IQ and age range. The DSM-IV added childhood disintegrative disorder, Asperger’s disorder, and Rett’s disorder to the already existing diagnoses of the pervasive developmental disorder not otherwise specified (PPD-NOS) category. Rett’s disorder, in which symptoms include problems with language, coordination, and repetitive movements, was removed from this category because it was found to be linked to genetics. Other disorders under this category are not thought to be genetic in origin at this time.

There were issues at this time classifying Asperger’s in the same category as autism. Professionals felt that the disorders may not be the same and should therefore not be under the same diagnosis in the DSM-5. [13] questioned the assumption that Asperger’s disorder and high functioning autism are quantitative manifestations of the same disorder [20]. They tested six male patients ages 6-12 on their visual cognitive function and behavior. The study shows that Asperger’s disorder and high func-
tioning autism are qualitatively different disorders. The children with autism had higher numbers of symptoms of anxiety and depression that was attributed to their discomfort in social situations.

During the revision process, the researchers and clinicians use the recent discoveries and scientific advances to consider what changes should be made to the current edition. Occasionally, drastic changes need to be made to the DSM because of such profound advances in research findings. The preparation of the DSM-5 encouraged almost 400 scientists from around the world to produce many new peer-reviewed articles.

The APA had more than 160 top researchers to help revise the DSM-IV and create the DSM-5. These professionals came from all different backgrounds including neuroscience, biology, genetics, statistics, epidemiology, social and behavioral sciences, nosology, and public health [6].

Until 2013, many developmental disabilities were considered separately for diagnosis. The DSM-5 contains some of the biggest changes made since the DSM’s first release. This newest edition combined autism, which is a childhood disorder characterized by significant impairment in social interactions and communication and by restricted patterns of behavior, interest, and activities, with Asperger’s disorder, childhood disintegrative disorder, and pervasive developmental disorder into one diagnosis, called Autism Spectrum Disorder [5]. ASD can range from mild to severe. The diagnosis is characterized by two groups of features. The first is persistent impairment in reciprocal social communication and social interaction and the second is restricted, repetitive patterns of behavior. Both of these would need to be present in early childhood to receive the diagnosis. Along with combining these disorders into one category, the separate diagnosis of Asperger’s was removed, which was the most drastic change to this section. Asperger’s disorder involves a significant impairment in the ability to
engage in meaningful social interactions. It is also accompanied with restrictive and repetitive stereotyped behaviors but without the severe delays in language or other cognitive skills characteristic of individuals with autism [3]. Even though the DSM is very thorough in its description and aids in the assessment and diagnosis of mental disorders, there is actually nothing in the manual that provides any guidance on the treatment or intervention that should be implemented in a patient’s plan.

While there is no cure for ASD, there are treatments available that have been proven effective. Occupational therapy, speech therapy, physical therapy are some examples of effective treatments for some characteristics of autism. Applied behavior analysis (ABA) is another one of those treatments.

ABA is a form of treatment where a behavioral interventionist will apply principles of learning and motivation from behavior analysis. Behavior analysis is the scientific study of behavior [1]. ABA is a widely used technique across many fields. It can be used to treat individuals with mental illnesses, developmental disabilities and learning disorders as well as improving skills in schools, homes, institutions, hospitals, and business centers. A focus of ABA therapy is to help effect behaviors such as language, social, academic, leisure and function life skills, aggression, self-injury, oppositional, and stereotyped behavior. ABA has been used as a treatment for autism since 1970 [14]. The goal of ABA treatment in individuals with autism is to find solutions to problems of social significance.

An ABA therapist, or behavioral interventionist is a professional who is board certified in behavior analysis (BCBA). Behavioral interventionists create a plan for the individual to be carried out by teachers and parents and others in order to help the individual diagnosed with ASD to succeed.

In ABA, a behavioral interventionist will systematically use interventions based on
the principles of learning theory to improve these behaviors. This demonstrates that interventions utilizing ABA strategies are likely responsible for the improvement in the individual’s behavior [1].

The interventionists at the Center for Autism and Related Disorders (CARD) use teaching approaches based on the well-established principles of ABA. Early Intensive Behavioral Intervention (EIBI) is a type of ABA for very young children with an ASD, usually younger than five, often younger than three. The basic goal in ABA treatment for autism is to bring about meaningful and positive change in behavior. There are several ways to go about bringing these changes.

One way ABA is used in autism treatment is to increase behaviors. This can be done by positive reinforcement. Positive reinforcement is when a behavior is followed by some sort of reward, making the behavior more likely to be repeated. Another method is to use reinforcement to teach life, communication, and social skills. In ABA, the goal is to not only change behaviors but also to maintain behaviors. A behavioral interventionist will also teach self-control and self-monitoring procedures to maintain and generalize skills. ABA therapy is not always done in the classroom or at home, so the behavioral interventionist also needs to teach the individual how to generalize, or transfer, the behavior from one situation or response to another. An example of this would be when the interventionist teaches the individual to complete the assigned work from class at home after the individual has mastered completing the assignment in a behavior session. ABA also aims to restrict or narrow conditions under which interfering behaviors occur (e.g., modifying the learning environment). Along with increasing some behaviors, an interventionist aims to reduce behaviors that could interfere with treatment success. Examples of these behaviors are self-injury, aggression, and stereotypy [1].

As the definition of ASD changes with each new edition of the DSM, the treatment
and understanding of ASD must change with it. Despite the years of research that were done in the area of ASD, it may still be unreasonable to think we can treat all these individuals the same if we initially categorized these disorders separately. This leads to our analysis of the phenotypes and how we aim to see how challenging behaviors are impacting learning outcomes in individuals diagnosed with ASD.

1.2 Thesis Guideline References

In this thesis we consider how the different phenotypes of ASD respond to applied behavior analysis treatment (hours vs exemplars mastered). The remainder of this paper is organized as follows: Related work is described in Chapter 2. Chapter 3 gives a breakdown of the dataset provided. Chapter 4 provides a description of the methods used in this paper. Chapter 5 details the results of the methods from Chapter 4. In Chapter 6 we discuss the meaning of the results found and their implications of ABA treatment for the separate phenotypes. Chapter 7 discusses future work to be done with the data. In Chapter 8, we discuss the conclusions from the study.
Chapter 2

Related Work

2.1 Treatment Intensity and Duration

In 2017, the authors in [12] performed an analysis of the effects of intensity and duration of treatment on eight separate skills for 1,468 children, ages 18 months to 12 years old, with autism spectrum disorder. Treatment intensity refers to the hours per week treatment is enforced. Treatment duration refers to how many months of treatment the patient experienced. All of these children were receiving ABA treatment.

<table>
<thead>
<tr>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
</tr>
<tr>
<td>Adaptive</td>
</tr>
<tr>
<td>Cognitive</td>
</tr>
<tr>
<td>Executive</td>
</tr>
<tr>
<td>Language</td>
</tr>
<tr>
<td>Motor</td>
</tr>
<tr>
<td>Play</td>
</tr>
<tr>
<td>Social</td>
</tr>
</tbody>
</table>

Table 2.1: Eight skills used by Linstead et al. in their multiple regression analysis

The authors used separate multiple linear regression analyses to study the relation-
ships between intensity and duration of treatment for all eight skills. A multiple linear regression model attempts to find the relationship between two or more predictor variables and a target variable by fitting a linear equation to the observed data.

The results from [12] suggested that treatment intensity and duration both were significant predictors of mastered learning objectives for all eight skills. The skills that showed the strongest response to treatment intensity and treatment duration were the academic and language domains.

2.2 Hierarchical Clustering on Patients Diagnosed with ASD

The authors in [17] had a sample of 2,400 children diagnosed with ASD. In [17] the authors used an unsupervised machine learning technique called hierarchical clustering to identify patterns of skills across eight developmental skills domains. These skills are the same as in [12]. Then a regression analysis was performed to examine the relationship between these phenotypes and learning outcomes. The authors in this paper wanted to be able to provide a solid foundation for more studies to be done on the relationship between ASD phenotypes and learning outcomes.

The goal of this study was to allow behavioral interventionists to provide personalized treatment and behavioral interventions for patients to maximize learning outcomes at the minimum time and cost.

This led to our analysis in this thesis of challenging behavior clusters. Our study further examines the relationship between therapy hours and exemplars mastered by
comparing the success of patients demonstrating different challenging behaviors. The goal of this work is to effectively treat the varying phenotypes of ASD.
Chapter 3

Data

3.1 The Center for Autism and Related Disorders

The data is from the Center for Autism and Related Disorders, commonly referred to as CARD. CARD utilizes the SKILLS\textsuperscript{TM} database. The SKILLS\textsuperscript{TM} database is a system provided by CARD and is a resource used for creating and implementing individualized treatment plans catered to the needs of each individual [16]. It is the responsibility of a BCBA to create an individualized plan for an individual diagnosed with ASD. Thanks to SKILLS\textsuperscript{TM}, each individual’s customized treatment plan can be accessed by clinicians, teachers, and parents so everyone involved can help the child succeed. SKILLS\textsuperscript{TM} lets the individual’s behaviors be tracked over extended periods of time using mobile applications and web-based software.

CARD provided a dataset of challenging behaviors for 2,116 patients and their gender. We were also provided records for 1,467 patients’ therapy hours and exemplars mastered. We joined these datasets based on their patient ids and we were left with 854 patients (148 females and 706 males).
For the patients in this study, the criteria for mastery is set by the behavioral interventionist. The interventionists at CARD require that the patients achieve greater than 70% accuracy of responding to the learning objective for a minimum of two treatment sessions across two different days [12]. Typically, 80% accuracy is needed to consider an exemplar mastered but interventionists are allowed to deviate from this criterion if in their professional opinion it is acceptable to do so [12].

The data consists of a patient id number for each individual, the challenging behavior rates, total therapy hours, total exemplars mastered, and gender.

The data has eight challenging behaviors and their associated frequencies for each patient. The frequencies were calculated by dividing the number of observations a challenging behavior was observed by the total number of challenging behaviors the patient demonstrated. This gave us a value between 0.0 and 1.0 for each challenging behavior per patient. The eight challenging behaviors are as detailed in Table 3.1.

<table>
<thead>
<tr>
<th>Challenging Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior</td>
</tr>
<tr>
<td>Aggression</td>
</tr>
<tr>
<td>Self-injury</td>
</tr>
<tr>
<td>Disruption</td>
</tr>
<tr>
<td>Elopement</td>
</tr>
<tr>
<td>Stereotypy</td>
</tr>
<tr>
<td>Tantrums</td>
</tr>
<tr>
<td>Non-compliance</td>
</tr>
<tr>
<td>Obsession</td>
</tr>
</tbody>
</table>

Table 3.1: Challenging Behaviors

3.2 Database

The data was stored in Microsoft SQL Server running on a 16-core Intel Xeon processor, 256GB of Ram, 256GB Solid State Hard drive and a 8TB spinning-disk hard
drive for data storage [17]. Data was extracted using Structured Query Language (SQL) in the form of various query statements.

Before we began our analyses on the data, we transformed the data by applying the logarithm function to the therapy hours and exemplars mastered using the R computing language. Using logarithmic scales helps us to take care of the larger values in our data set in order to scale down these larger variables.
Chapter 4

Methods

All methods were performed using the programming language R.

4.1 K-Means Clustering

The first algorithm we used on this data was k-means clustering. Clustering is an unsupervised machine learning technique that can find meaningful relationships in data where labelling information is either not available upfront, or simply not present in the data. The goal of clustering is to find natural groups, or clusters, in data. Data within the same cluster will have more similar features than data within different clusters [15].

K-means is a widely used prototype-based clustering algorithm. This means that each cluster is represented by a prototype. This prototype can either be the centroid of data points with similar continuous features, or the medoid if we are dealing with categorical features. For our algorithm, we are looking at continuous features, so we have a centroid for each cluster.
We use k-means here as our method for clustering because compared to other clustering methods it is computationally efficient and easily implemented.

The k-means technique has these 5 steps:

1) Choose the number of clusters, k. As easy and computationally efficient as k-means is, it is not usually obvious what the value of k should be. To find the best value of k, we incrementally tested values of k between 2 and 20.

2) For each of these values, the algorithm picks k sample points from the data at random. These are the initial centroids \((c_1, c_2, ..., c_k)\).

3) Then for each data point \(d_i\), it assigns \(d_i\) to the nearest centroid \(c_k\).

4) After this, the algorithm recalculates the centroids.

5) The algorithm repeats steps 3 and 4 until the cluster assignments do not change or a maximum number of iterations is reached.

When using k-means with data that has continuous features, we use squared Euclidean distance to find the distance between the data points and the centroids.

A question we need to answer now is how to measure the similarity between data points. We define similarity as the opposite of distance. A commonly used metric for finding the distance between two data points \(x\) and \(y\) in \(m\)-dimensional space is the squared Euclidean distance:

\[
d(x, y)^2 = \sum_{j=1}^{m} (x_j - y_j)^2 = ||x - y||_2^2
\]  

From these similarities, we can turn clustering into an optimization problem. We can take an iterative approach for minimizing the within-cluster Sum of Squared Errors.
(SSE), or cluster inertia. Once these errors are calculated, we can look at a graph of the errors and by using the elbow method we can decide the best value for k. The elbow method is a method in which we can look at the line of the graph as the ”arm” and the ”elbow” is the point in which we start to see diminishing returns. As k increases, our SSE gets smaller. When k equals the number of points in our dataset, the SSE is 0, and every point is its own cluster.

$$SSE = \sum_{i=1}^{m} \sum_{j=1}^{k} w_{i,j} ||d_i - c_j||_2^2 \quad (4.2)$$

Another metric commonly used to solidify confidence in a clustering algorithm is silhouette analysis. Using silhouette graphical tool in R, we can plot and see how tightly grouped the samples are in the clusters. In 3 steps, we can also calculate the silhouette coefficient [15].

1. Calculate how closely related objects in a cluster $$a^i$$ are (cluster cohesion) as the average distance between a sample $$x^i$$ and all other points in the same cluster
2. Calculate the cluster separation $$b^i$$ from the cluster closest to that cluster as the average distance between the sample $$x^i$$ and all samples in the nearest cluster
3. Calculate the silhouette $$s^i$$ as the difference between cluster cohesion and separation divided by the greater of the two

$$s^i = \frac{b^i - a^i}{\max(b^i, a^i)} \quad (4.3)$$

Where $$s^i$$ is between -1 and 1. If the silhouette coefficient is 0 then the cluster separation and cohesion are equal to each other ($$b^i = a^i$$).
Our ideal silhouette coefficient is 1, and this is attained when $b_i$ is significantly greater than $a_i$.

### 4.2 Linear Regression

Supervised learning is when we have labels for our data and our algorithm learns from these labels. For example, if we wanted to predict the exemplars mastered by a patient in advance, we can get the therapy hours of the patients we have observed and have a supervised learning algorithm predict how to associate the exemplars mastered to the patient.

The authors in [12] used a multiple linear regression model to evaluate the relationship between treatment intensity and treatment duration on the mastery of skills. This is what we will use in this paper as well. In our study, our explanatory variables are therapy hours, cluster, and gender and our target variable is exemplars mastered.

In simple linear regression, otherwise known as univariate linear regression, we are trying to model the relationship between a single explanatory variable $x$ and a response, or target, variable $y$. The equation used for linear models with only one predictor variable is defined as follows:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$ (4.4)

In this equation, the weight $\beta_0$ represents the y axis intercepts and $\beta_1$ is the coefficient of the explanatory variable. In simple linear regression we want to find the weights...
of the equation to explain the relationship between the explanatory variable and the target variable. From this, we can also predict the responses of new data points that were not part of the observed data.

We can generalize equation 4.4 to get an equation for multiple linear regression where we have multiple variables (equation 4.5).

\[ y_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 w_i + \epsilon_i \] (4.5)

Linear regression works by taking the explanatory variables \( (u, v, \text{ and } w) \) and the response variable \( y \) and fitting a straight line to the data that minimizes the distance between our observed point and the fitted line. The line-of-best-fit is commonly referred to as the regression line. In our study, our explanatory variables are treatment hours, cluster, and gender and our response variable is exemplars mastered.

A good way to quantitatively measure a model’s performance is the Mean Squared Error (MSE). The MSE is the average of the value of the SSE cost function that is minimized to fit the linear regression model. We can use the MSE to compare different regression models.

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y^i - \hat{y}^i)^2 \] (4.6)
\( R^2 \) is the standardized version of the MSE used for better interpretability of the model’s performance. \( R^2 \) represents the response variance captured by the model.

\[
R^2 = 1 - \frac{SSE}{SST} \quad (4.7)
\]

where SSE is the sums of squared errors and SST is:

\[
SST = \sum_{i=1}^{n} (y^i - \mu_y)^2 \quad (4.8)
\]

It is easy to see that \( R^2 \) is a rescaled version of the MSE (equation 4.9).

\[
R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{1}{n} \sum_{i=1}^{n} (y^i - \hat{y}^i)^2 = 1 - \frac{MSE}{Var(y)} \quad (4.9)
\]

\( R^2 \) is bounded between 0 and 1. If \( R^2 = 1 \) then we have a perfect relationship between \( x \) and \( y \) and our \( MSE = 0 \).

\( R \) allows us to specify interactions terms in our regression formulas. An interaction occurs when the product of two predictor variables is also a significant predictor [18].

We have three explanatory variables (therapy hours, cluster, and gender) and we want to include all their interactions in our model.

In R, we write this as \( y \sim u * v * w \).

Incorporating these three explanatory variables into our equations gives us equation 4.10.

\[
y_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 w_i + \beta_4 u_i v_i + \beta_5 u_i w_i + \beta_6 v_i w_i + \beta_7 u_i v_i w_i + \epsilon_i \quad (4.10)
\]
Chapter 5

Results

5.1 Clustering Results

Figure 5.1 shows the within-groups sums of squared differences for the female patients. We can see that 7 is the best value of k for females since this is where we see diminishing returns on the graph.

Figure 5.1: Sums of Squared Differences for Females
Figure 5.2 shows the within-groups sums of squared differences for the male patients. We can see that 7 is the best value of k for males since this is where we see diminishing returns on the graph.

The silhouette plots in Figures 5.3 for each gender also indicate that 7 is a good choice for females and males. The silhouette plot with all patients in Figure 5.4 confidently reaffirms that 7 is a good choice for k for all patients with an average silhouette width of 0.65. So, for our k-means algorithm we picked a value of 7 for k.
Figure 5.4: Silhouette Plot for All Patients

As we look at our silhouette plots, we observe that our silhouette coefficients are very far away from 0. This is an indication that our number of clusters is appropriate for the data.

After running the k-means algorithm, we calculated the average frequency of all eight challenging behaviors for each cluster and found seven phenotypes of ASD, most of which demonstrate one dominant challenging behavior. Table 5.1 shows how many individuals are in each cluster and the dominant challenging behavior of the cluster.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Challenging Behavior</th>
<th>Females</th>
<th>Males</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tantrums</td>
<td>14</td>
<td>60</td>
<td>74</td>
</tr>
<tr>
<td>2</td>
<td>self-injurious</td>
<td>8</td>
<td>79</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>elopement</td>
<td>18</td>
<td>78</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>stereotypy (low rate)</td>
<td>37</td>
<td>170</td>
<td>207</td>
</tr>
<tr>
<td>5</td>
<td>non-compliance</td>
<td>26</td>
<td>87</td>
<td>113</td>
</tr>
<tr>
<td>6</td>
<td>aggression</td>
<td>26</td>
<td>113</td>
<td>139</td>
</tr>
<tr>
<td>7</td>
<td>stereotypy (high rate)</td>
<td>19</td>
<td>119</td>
<td>138</td>
</tr>
</tbody>
</table>

Table 5.1: This table shows the breakdown of how many individuals are in each cluster and the dominant challenging behavior of the phenotype.
This table is not the best visual representation of the phenotypes, so we created a radar chart to best display the eight challenging behaviors in our data. We display these phenotypes in Figure 5.5 and Figure 5.6. Figure 5.5 shows how each cluster differs from the others in terms of the challenging behaviors. Figure 5.6 shows individual graphs of each cluster to better show the challenging behavior dominant for each cluster. The radar charts in Figure 5.6 are scaled from 0 to the average frequency of the dominant challenging behavior. For example, Cluster 1 is scaled from 0 to 0.6, where tantrums extends to. Cluster 4 is scaled from 0 to 0.4, where stereotypy extends to.

It is worth noting that cluster 4 and cluster 7 both have stereotypy as their dominant challenging behavior, but at different frequencies (Figure 5.7). We comment on this similarity in the following chapter.
Figure 5.6: Radar Charts for Phenotypes
5.2 Linear Regression Results

$R^2$ is a measure of our model’s quality. Our model’s $R^2$ value was 0.6662. Remember that the value for $R^2$ is the fraction of the variance of the exemplars mastered that is explained by our model. Our model explained 66.62% of the variance of exemplars mastered, and the remaining 33.38% is left unexplained.

The F-statistic is an indication of whether our model is significant or insignificant. If any of the coefficients are nonzero, then the model is significant. The model is insignificant if all of the coefficients are 0. We have nonzero coefficients, so the F-statistic says our model is significant (F-statistic: 261.05 on 27 and 826 DF).

Figure 5.8 shows the regression lines for all the different clusters. We can take a look at the MSE for each cluster, shown in Table 5.2. Figure 5.9 shows the regression lines for the separate genders. Figure 5.10 separates each gender by cluster.

Figure 5.11 shows a box plot of the 7 clusters. The box plots show the range of the exemplars mastered for each cluster, where the whiskers represent the minimum and maximum values. The line across each box is the mean. The top of the box represents the third quartile. The bottom of the box represents the first quartile. Any points on graph represent outliers in the clusters. Here we can start to examine the differences
Figure 5.8: Line-of-Best Fit for Each Cluster

![Line-of-Best Fit for Each Cluster](image)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Challenging Behavior</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tantrums</td>
<td>0.2981379</td>
</tr>
<tr>
<td>2</td>
<td>self-injurious</td>
<td>0.3390162</td>
</tr>
<tr>
<td>3</td>
<td>elopement</td>
<td>0.2323062</td>
</tr>
<tr>
<td>4</td>
<td>stereotypy (low rate)</td>
<td>0.3717557</td>
</tr>
<tr>
<td>5</td>
<td>non-compliance</td>
<td>0.2954241</td>
</tr>
<tr>
<td>6</td>
<td>aggression</td>
<td>0.2440494</td>
</tr>
<tr>
<td>7</td>
<td>stereotypy (high rate)</td>
<td>0.2864331</td>
</tr>
</tbody>
</table>

Table 5.2: Mean Squared Error Comparison

in the clusters and exemplars mastered and start to analyze if clusters are performing differently.

The p-value is the probability that estimates the likelihood that the coefficient is not significant. If p-values are big, they are indicating that the corresponding variables are insignificant predictors. Variables that have large p-values are candidates for elimination from our model [18].
We can see all of the p-values from the model in Table 5.4. We have a significant p-value for therapy hours ($p-value < 2.2e-16$). This means that therapy hours has a significant relationship with exemplars mastered. Cluster also has a significant relationship with exemplars mastered ($p-value = 0.001503$). However, there is not a significant p-value for the interaction between therapy hours and cluster ($p-value = 0.276400$). Gender is not a significant either ($p-value = 0.050575$). However, the interaction between gender and cluster does have a significant relationship with exemplars mastered ($p-value = 0.017966$).

Table 5.3 shows us the averages for therapy hours and exemplars mastered for each cluster. We can see that cluster 4 has the highest mean out of all of the clusters. Cluster 3 has the second highest average number of exemplars mastered. Cluster 2 has the lowest average number of exemplars mastered.
We can look at the interaction between gender and cluster in Figure 5.12. Our ANOVA indicates that the genders perform significantly different for several of the clusters even though they exhibit the same challenging behaviors. Table 5.5 provides the average therapy hours and exemplars mastered per cluster for females. Table 5.6 provides the average therapy hours and exemplars mastered per cluster for males.
The first step we needed to take was to assess whether there were any differences between our clusters and the exemplars mastered. Our ANOVA results indicated there was a difference between therapy hours, cluster, and the interaction of cluster and gender. Now, we can do a follow-up test, or post-hoc analysis. This will help
Figure 5.12: Boxplots of Each Cluster and Gender

Clusters of Female Patients

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Challenging Behavior</th>
<th>Hours</th>
<th>Exemplars Mastered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tantrums</td>
<td>6.487342</td>
<td>5.366978</td>
</tr>
<tr>
<td>2</td>
<td>self-injurious</td>
<td>6.720123</td>
<td>5.433814</td>
</tr>
<tr>
<td>3</td>
<td>elopement</td>
<td>6.664452</td>
<td>5.500867</td>
</tr>
<tr>
<td>4</td>
<td>stereotypy (low rate)</td>
<td>6.900855</td>
<td>5.800398</td>
</tr>
<tr>
<td>5</td>
<td>non-compliance</td>
<td>6.567966</td>
<td>5.390436</td>
</tr>
<tr>
<td>6</td>
<td>aggression</td>
<td>6.579717</td>
<td>5.147397</td>
</tr>
<tr>
<td>7</td>
<td>stereotypy (high rate)</td>
<td>6.656241</td>
<td>5.131328</td>
</tr>
</tbody>
</table>

Table 5.5: Averages of therapy hours and exemplars mastered for the females of each cluster

us determine which of the clusters differ and estimate by how much they differ. We used the Tukey method to do our multiple comparisons.

The Tukey post-hoc test compares groups after we have already observed a significant p-value. Tukey HSD is more conservative because it is a protected t-test which means the alpha value is lower and it is harder to reach statistical significance.
Clusters of Male Patients

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Challenging Behavior</th>
<th>Hours</th>
<th>Exemplars Mastered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tantrums</td>
<td>6.606496</td>
<td>5.331677</td>
</tr>
<tr>
<td>2</td>
<td>self-injurious</td>
<td>6.638475</td>
<td>5.045405</td>
</tr>
<tr>
<td>3</td>
<td>elopement</td>
<td>6.706360</td>
<td>5.410371</td>
</tr>
<tr>
<td>4</td>
<td>stereotypy (low rate)</td>
<td>6.917946</td>
<td>5.496782</td>
</tr>
<tr>
<td>5</td>
<td>non-compliance</td>
<td>6.474624</td>
<td>5.258923</td>
</tr>
<tr>
<td>6</td>
<td>aggression</td>
<td>6.538419</td>
<td>5.265986</td>
</tr>
<tr>
<td>7</td>
<td>stereotypy (high rate)</td>
<td>6.722390</td>
<td>5.413940</td>
</tr>
</tbody>
</table>

Table 5.6: Averages of therapy hours and exemplars mastered for the males of each cluster

The Tukey post-hoc test tests all the pairwise comparisons using the Tukey HSD (honest significant difference) formula displayed in equation 5.1.

\[
HSD = \frac{M_i - M_j}{\sqrt{\frac{MS_W}{n_h}}}
\]  

(5.1)

where \(M_i - M_j\) is the difference between the pairs of means, \(MS_W\) is the mean square within and \(n\) is the number of clusters.

By looking at Figure 5.13, we can see that we have two pairwise tests that show a significant difference. We have a significant difference between cluster 4 (low frequency stereotypy and moderate frequencies of other challenging behaviors) and cluster 2 (self-injurious) as indicated by the p-value \((p-value = 0.0025954)\) and between cluster 4 and cluster 6 (aggression) \((p-value = 0.0472675)\). We know that cluster 4 is our overall highest mastery cluster and that cluster 2 is the lowest mastery cluster, but our post-hoc results confirm that there is a significant difference between the clusters.
We can plot the multiple comparisons and construct a confidence interval for each pairwise difference in the means. We have 7 confidence intervals in this case. We can find the significantly different clusters in Figure 5.13 by observing the confidence intervals where the interval does not contain 0.

Figure 5.13: Tukey Post Hoc for Clusters

Our model returned a significant p-value for the interaction between cluster and gender, so we also want to see where the differences between these two variables. Figure 5.14 shows the confidence intervals for the gender and cluster interaction. This figure has many confidence intervals and it is difficult to see which intervals do not contain 0. Luckily, we can extract the p-values from the post-hoc analysis. We found that the females and males in cluster 4 significantly mastered more exemplars than the males in cluster 2 with \( p-value = 0.0050173 \) and \( p-value = 0.0311176 \), respectively. There was no significant difference between the females in cluster 2 and
the females and males in cluster 4. There were also no significant differences within clusters between the genders.

Figure 5.14: Tukey Post Hoc for Gender*Clusters
Chapter 6

Discussion

From the results, we see that low frequencies of stereotypy and moderate frequencies of other challenging behaviors in cluster 4 do not appear to interfere with ABA treatment as significantly as other challenging behaviors like non-compliance and self-injurious. Cluster 4 seems to have stereotypy as the dominant behavior, but there are moderate frequencies of other challenging behaviors. This could be due to different types of stereotypic behaviors being demonstrated by the patients in cluster 4 and 7.

Cluster 6 is the cluster most afflicted by aggressive behavior, and we can see that it significantly negatively affects the exemplars mastered by individuals in this group.

Self-injurious behavior also negatively affects the exemplars mastered by the individuals in cluster 2. SIB is known to be the most devastating challenging behavior to caregivers and providers. SIB can range from mild head rubbing up to severe head banging and can become life threatening [7]. There are many theories that impaired cognitive ability in individuals with ASD causes SIB [4]. A study done by [13] found that poor receptive and expressive communication was associated with higher levels of SIB. We can see that this reflects in the results of our data. We have the SIB
cluster mastering significantly less exemplars than cluster 4 where we see almost no frequency for SIB. Patients who have higher frequencies have a difficult time communicating, so they resort to self-injurious behavior. A study by the authors in [8] found that age and SIB are negatively correlated, meaning that SIB declines as the patient gets older.

[10] found that the challenging behaviors to caregivers (unusual vocalizations and play with objects, aggression towards others) were related with ASD severity. The authors in [7] hypothesized that SIB or aggression directed towards the self may also be related to severity of ASD. Since cluster 2 demonstrated SIB at a high frequency, perhaps a different approach for treatment should be considered. The same goes with cluster 6. Aggression may be interfering with treatment to the point that the treatment is not effective.
Chapter 7

Future Work

An interesting future study for this dataset could be to look at the average age of patients within each clusters. The authors in [8] found that age and SIB were negatively correlated. It would be interesting to see if the SIB cluster had the lowest average age.

Additional work stemming from this study could include mapping function of behavior onto the phenotypes we found. In ABA, a behavioral interventionist is aiming to replace challenging behaviors with functionally equivalent pro-social behaviors. In other words, the ABA therapist is trying to teach a new behavior that will result in the same outcome for the patient without needing to resort to a challenging behavior. For example, hitting results in the interventionist leaving the room, but verbally communicating ”leave me alone” results the same outcome. Therefore, teaching the phrase ”leave me alone” would reduce and eventually replace hitting.
Chapter 8

Conclusion

In this thesis, we looked at the data provided by CARD for 854 individuals diagnosed with ASD and receiving Applied Behavior Analysis treatment. We found 7 phenotypes of ASD using k-means clustering. Each phenotype we found had a dominant challenging behavior.

After finding these phenotypes, a multiple linear regression analysis was performed on the data to explore the relationship between exemplars mastered and therapy hours, cluster, and gender. We looked at the interaction between all of the explanatory variables as well. We found that therapy hours are a significant variable in this study. Cluster is also a significant variable. The only interaction that gave us significant results was cluster and gender.

We then used Tukey’s post-hoc test to find the pairs that were significantly different in our data. When we looked into the cluster differences, we found that cluster 4 (low stereotypy and moderate frequencies of other challenging behaviors) was significantly more successful at mastering exemplars than both cluster 2 (self-injurious) and cluster 6 (aggression). We found that the females in cluster 4 were significantly more
successful than the males in cluster 2, and the males in cluster 4 were significantly more successful than the males in cluster 2. We did not find that any gender had more success than the other within the same cluster.

The conclusion of this study is intended to help inform behavioral interventionists of how challenging behaviors impact ABA treatment for children diagnosed with ASD. The work we have presented will help behavioral interventionists create personalized therapeutic interventions with maximum efficacy, minimum time and minimum cost for individuals.
References


## Appendices

### A Supplemental Materials

#### Figure A.1: Example of Patient Data

<table>
<thead>
<tr>
<th>Discontinuation Cause</th>
<th>Discontinuation, n (%)</th>
<th>Discontinuation, n (%)</th>
<th>Discontinuation, n (%)</th>
<th>Discontinuation, n (%)</th>
<th>Discontinuation, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>Drug-related</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>Adverse event</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>Other</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>Withdrawal</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>AE</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>AE-related</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>AE-unknown</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>AE-related</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
<tr>
<td>AE-unknown</td>
<td>292</td>
<td>574</td>
<td>352</td>
<td>352</td>
<td>352</td>
</tr>
</tbody>
</table>

41