Running head: ASD Data Analysis

# Autism Spectrum Disorder (ASD) Data Analysis

# Patrick Weatherford

# Bellevue University – DSC 530 Data Exploration and Analysis

# Abstract

Exploratory Data Analysis (EDA) was performed to find predictors of Autism Spectrum Disorder prevalence (ASD) in the United States. All predictors analyzed were aggregated and filtered by the 50 contiguous U.S. States. Ultimately the 2 features that were found to have an effect were ‘Vaccination Rate’ which is an average of the vaccination rate per State and ‘Father Age >= 40 Rate’ which is the rate per State at which fathers were greater than or equal to 40 years of age at their child’s birth. After linear regression analysis was ran however, the predictive model was not found to predict ASD rates well. The model itself having an adjusted *R*2 = 0.05, Vaccination Rate [p > 0.05, CI (-0.01, 0.14)], and Father Age >= 40 Rate [p > 0.05, CI (-8.76, 3.22)].

# Autism Spectrum Disorder (ASD) Data Analysis

“Autism spectrum disorder (ASD) is a developmental disability that can cause significant social, communication and behavioral challenges. People with ASD may communicate, interact, behave, and learn in ways that are different from most other people” (CDC). When diagnosing a patient, 2 criteria are used to determine how severe the disorder is: 1) Social impairment, 2) Degree of restricted and/or repetitive behaviors. Based on these 2 criteria, the level of severity is determined and can range anywhere from low-functioning to high-functioning.

In the United States, the prevalence of ASD as of 2018 was estimated to be 23 per 1000 (2.3%). That equates to roughly 1 out of every 44 people. The prevalence has been steadily increasing over time as well. For example, in 2000 the prevalence was 0.7% and in 2011 the prevalence was 1.1%. The specific cause of ASD is unclear but there have been specific gene expressions and environmental risk factors that have been found to increase the risk of ASD prevalence. Also, there are no known cures for ASD, but behavior intervention is often used for low-functioning patients in hopes that they become independent.

For this analysis, data exploration was conducted on environmental risk factors aggregated by U.S. State and then analyzed using various statistical methods to determine if they are potential predictors of ASD prevalence. Potential predictors will then be further analyzed to determine if they are significant and indicate an effect and if so, how much of an effect exists.

# Obtaining the Data

Initially the goal was to obtain a cross-sectional data set at a particular point in time, but this ended up being more challenging than expected. For example, many of the morbidity related surveys and studies were performed in cycles greater than 1 year or were pre-aggregated over multiple years. Ultimately, a Meta-analysis data set was obtained which combines data from multiple different surveys and studies and aggregates them together.

Missed Items During Analysis

There are some areas in the overall analysis where I could have gone more in-depth but overall, I would say that the main concepts of Exploratory Data Analysis (EDA) were examined. Some items that could have probably been improved below.

*Data*

More observations would have been beneficial for my analysis but as mentioned, find the data was rather challenging. Also, more predictors to test and analyze would have beneficial. With ASD rates increasing year-over-year, I am confident that there are environmental factors causing this increase. In future analysis I would dig deeper into time periods where the ASD rate increased rapidly and do more research on other predictors during this time. Lesson learned: devote more time to planning and obtaining data for your research.

*Analysis with Different Models*

Linear regression is a model that still works relatively well and has stood the test of time, but it is still relatively simple compared to other more advanced models that may have been better to use.

# References

# Basics About Autism Spectrum Disorder (ASD) | NCBDDD | CDC. (2020, August 11). Retrieved from <https://www.cdc.gov/ncbddd/autism/facts.html>

# Child and Adolescent Health Measurement Initiative. 2018-2019 National Survey of Children’s Health (NSCH) data query. Data Resource Center for Child and Adolescent Health supported by the U.S. Department of Health and Human Services, Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB). Retrieved 12/2/2021 from [www.childhealthdata.org](http://www.childhealthdata.org)

# Explore Census Data. (2021). Retrieved 2 December 2021, from https://data.census.gov/cedsci/

# Vaccination Coverage among Young Children (0-35 Months) | Data | Centers for Disease Control and Prevention. (2021). Retrieved 2 December 2021, from <https://data.cdc.gov/Child-Vaccinations/Vaccination-Coverage-among-Young-Children-0-35-Mon/fhky-rtsk>

# Centers for Disease Control and Prevention. CDC Wonder. <http://wonder.cdc.gov/>. December 2021.

# CDC BRFSS. (2022). Retrieved 17 January 2022, from <https://www.cdc.gov/brfss/>

# CDC BRFSS. (2022). Retrieved 20 January 2022, from <https://www.cdc.gov/brfss/>