Milestone2-example

Matthew Connell, Firas Moosvi

05/03/2020

Autism Screening

Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that impairs social interpretation/communication ability, as well as the presence of repetitive behaviors. Current diagnostic procedures are lengthy and inefficient (Thabtah 2019). Affecting 1.5% of the population, with many more cases going undetected, an easy-to-implement, effective screening method is warranted. ASDTest, a mobile app, has been introduced to provide an accessible screening method that tells the user whether they should seek professional healthcare opinions, based on a 10 question survey (Allison, Auyeung, and Baron-Cohen 2012). The ability to recognize and diagnose ASD at an early age can allow the affected to access the healthcare resources and support they will need, in a timely manner.

The Autism Spectrum Quotient-10 (AQ-10) consists of 10 questions intended to differentiate characteristics of autism in individuals. Each question has four possible answers: "Definitely Agree", "Slightly Agree,"Slightly Disagree", and"Definitely Disagree". For questions 1, 5, 7, and 10, a value of 1 is assigned for either a"slightly agree" or a "definitely agree" response. For questions 2, 3, 4, 6, 8, and 9, a value of 1 is assigned for either a "slightly disagree" or a "definitely. disagree" response. A cumulative score is calculated and a participant who receives a total score of greater than 6 is recommended for a specialist diagnostic assessment.

Data Description

The dataset used in this analysis was obtained from the University of California Irvine Machine learning Repository, uploaded by Fadi Thabtah. Each row represents an individual who participated in the survey. The survey's results, the app's classification, and some background information about the individual was recorded. Below is the entire variable set:

Variable

Type

Description

A1_score

Int (0,1)

Prompt: I often notice small sounds when others do not

A2_score

Int (0,1)

Prompt: I usually concentrate more on the whole picture, rather than the small details

A3_score

Int (0,1)

Prompt: I find it easy to do more than one thing at once

A4_score

Int (0,1)

Prompt: If there is an interruption, I can switch back to what I was doing very quickly

A5_score

Int (0,1)

Prompt: I find it easy to 'read between the lines' when someone is talking to me

A6_score

Int (0,1)

Prompt: I know how to tell if someone listening to me is getting bored

A7_score

Int (0,1)

Prompt: When I'm reading a story I find it difficult to work out the characters' intentions

A8_score

Int (0,1)

Prompt: I like to collect information about categories of things(e.g. types of car, types of bird, types of train, types of plant etc)

A9_score

Int (0,1)

Prompt: I find it easy to work out what someone is thinking or feeling just by looking at their face

A10_score

Int (0,1)

Prompt: I find it difficult to work out people's intentions

Age

Int

Age of the individual

Gender

String

M (male) or F (female)

Ethnicity

String

Common Ethnicities defined for each individual

Born with Jaundice?

String (yes,no)

Was individual born with jaundice?

Country of Residence

String

Home country of individual

Used app before?

String (yes, no)

Has the user has used a screening app

Result

Int

Cumulative score of the 10 survey Q's

age_desc

String

Age Group

relation

String

Parent, self, caregiver, medical staff, clinician ,etc.

ASD/Class

String (yes, no)

App's classification based on result
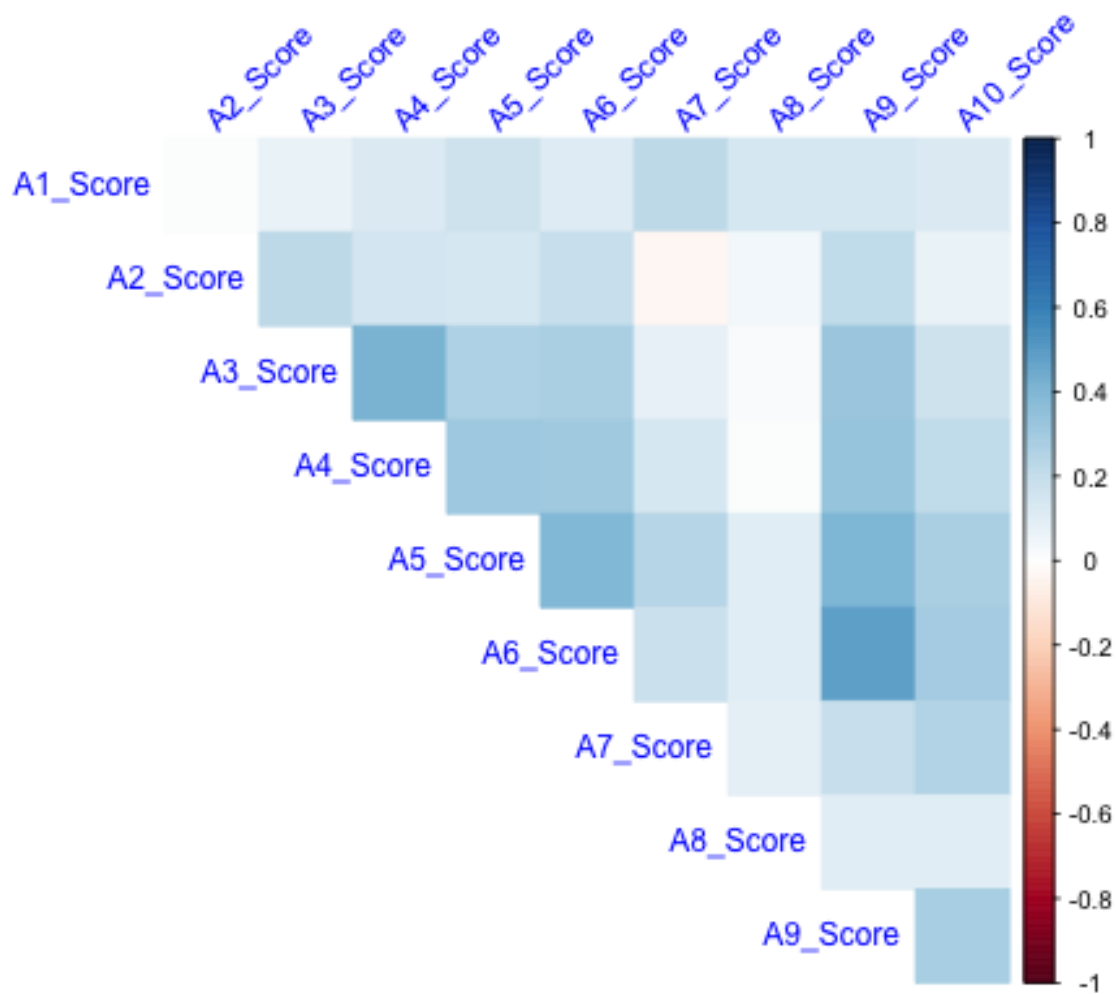
autism (Target Variable)

String (yes, no)

Does individual have an autism diagnosis?

Exploring the Dataset

Correllogram

From the correllogram below, we can see that there's very little correlation among any of the ten 'question' variables in the dataset. The colour scheme shows all positive correlations as blue, and all negative correlations as red.

Dodged Bar Chart

The dodged bar chart below shows the occurences of autism in people of different ethnicities who took the survey. This plot also illustrates an issue with the dataset: there are two levels called 'others' and one called NA. We should definitely combine the two 'others' columns and also decide what to do about the NAs.

Occurence of Autism by Ethnicity
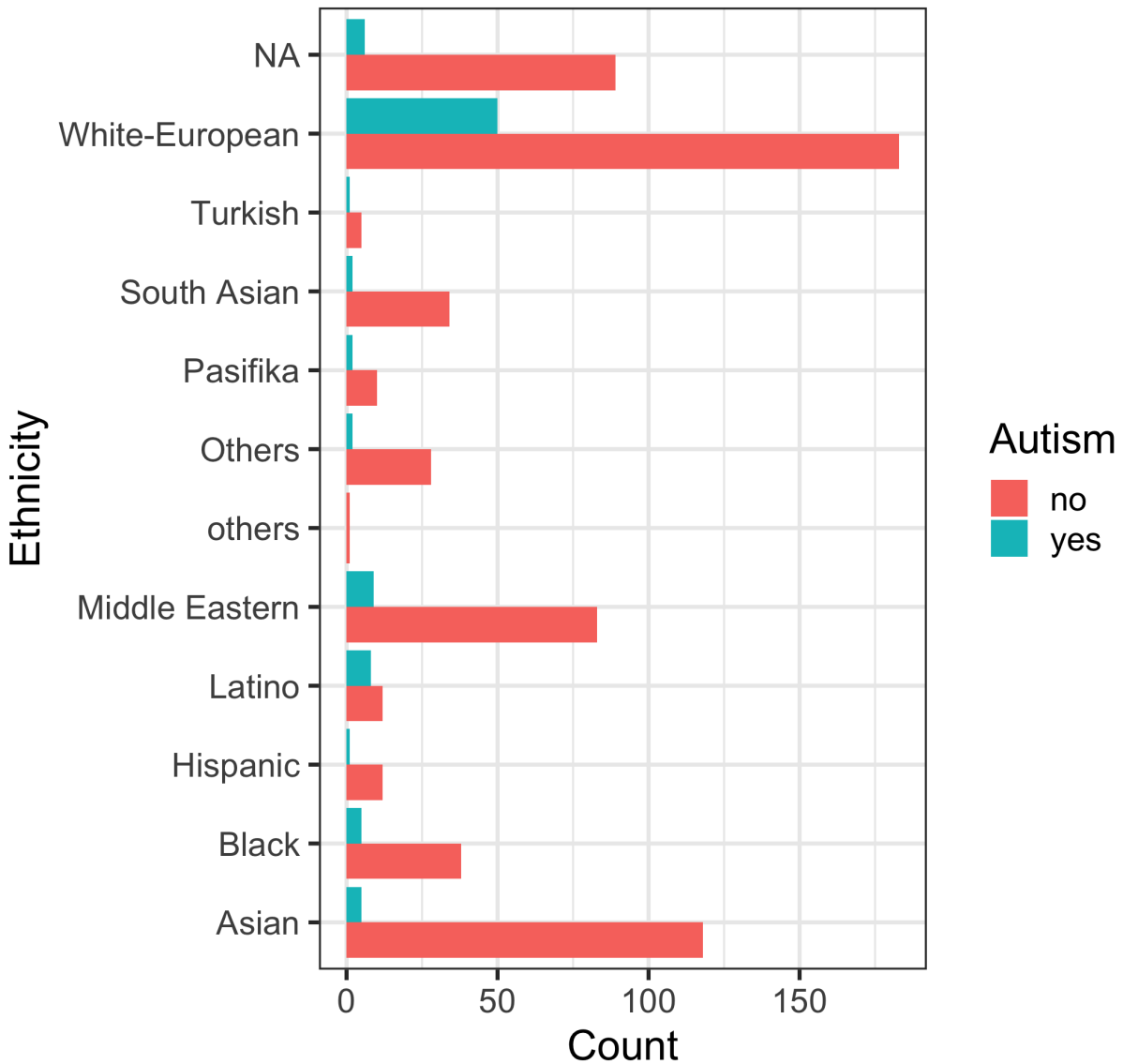
Proportional Bar Chart

The proportional bar chart below shows the percentage of people who were diagnosed with autism given a particular score on the autism screening test. A score of 0 would mean that it's incredibly unlikely the person would be diagnosed with autism. In general, the higher the score, the more likely it is that the person has autism.

Proportion of Participants by ASD-10 Test Score
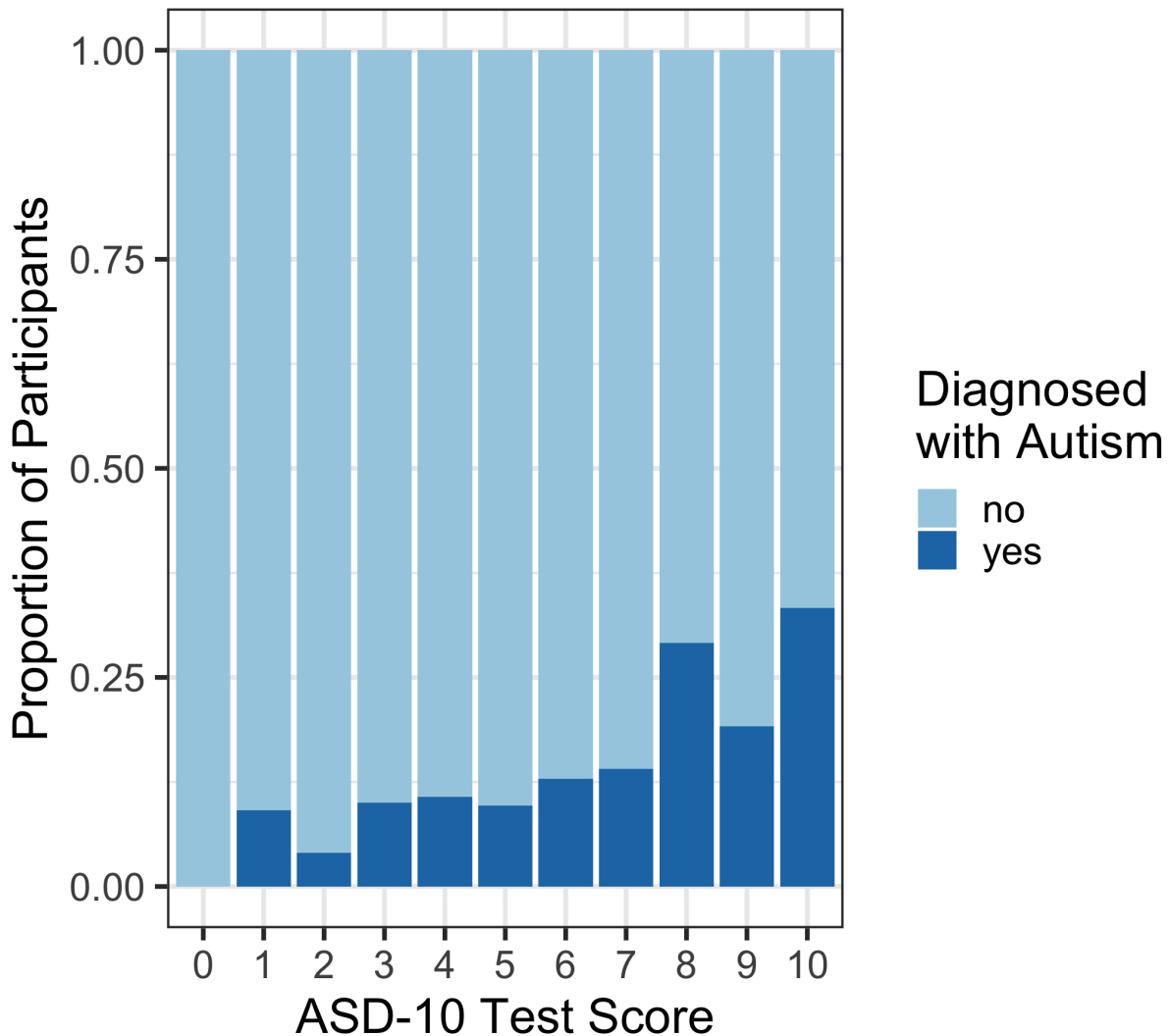
Research Question

In this analysis, we will use logistic regression to determine the relationship between the AQ-10 result (cumulative score) and the age of the individual.

Plan of Action

With our research question, we are only interested in the cumulative score and the age of the individual taking the survey. We will ignore the other variables for the purposes of this analysis. After dealing with the missing data, we will perform a linear regression analysis and plot the relevant variables with a regression line.

References

Thomas Pin and Tejas Phaterpekar, Matthew Connell - Autism Spectrum Disorder Screening Machine Learning Analysis - https://github.com/UBC-MDS/522-Workflows-Group-414

Allison, Carrie, Bonnie Auyeung, and Simon Baron-Cohen. 2012. "Toward Brief 'Red Flags' for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls." *Journal of the American Academy of Child & Adolescent Psychiatry* 51 (2). Elsevier: 202–12.

Thabtah, Fadi. 2019. "An Accessible and Efficient Autism Screening Method for Behavioural Data and Predictive Analyses." *Health Informatics Journal* 25 (4): 1739–55. https://doi.org/10.1177/1460458218796636.