Written Report on Factors that Help Diagnose Autism

The Repos - Jeffrey Bohrer, Alexandra Green, Anna Zhang, Kevin Lee

2025-03-20

I. Introduction

Autism Spectrum Disorder (ASD) remains a highly prevalent condition despite modern strides made in medical technology. It is reported that nearly 2.2% of adults are affected by ASD, and growing awareness has led to an uptick in diagnoses, particularly in adults who went undiagnosed early in life (Hirota 2023). However, ASD screening tests for all age groups currently contain significant inaccuracies. For example, the most widely used toddler screening test, CHAT-R/FAs, was recently found to produce false negatives in 25% of cases. In contrast, the most commonly used adult autism screening test – the Autism-Spectrum Quotient (AQ) – was found to have limited predictive value in certain populations (Aishworiya 2023; Curnow 2023). Therefore, it has become critical to identify stronger predictors or explore underlying relationships to have more accurate tests and models to predict ASD in adults.

In this study, we will focus on identifying the features that most greatly affect the probability of being encouraged to pursue a diagnosis within this particular questionaire, as created by Prof. Fadi Thabtau of the Manukau Institute of Technology. The data was sourced from users of his app, ASDTests, which screens its users for potential indicators of autism using a ten-question survey. The data set being used will contain ten characteristics along with the answers of each individual to this survey. Because ASD is difficult to identify and can significantly impair an individual's quality of life, understanding the relationship between demographics, certain behaviors, and their association with autism could encourage individuals to seek diagnosis and gain the self-understanding they need. These adults who receive a positive diagnosis can then access the necessary resources for support. Our goal is to identify which aspects of this questionnaire are most closely associated with being encouraged to pursue an autism diagnosis.

i. Univariate EDA

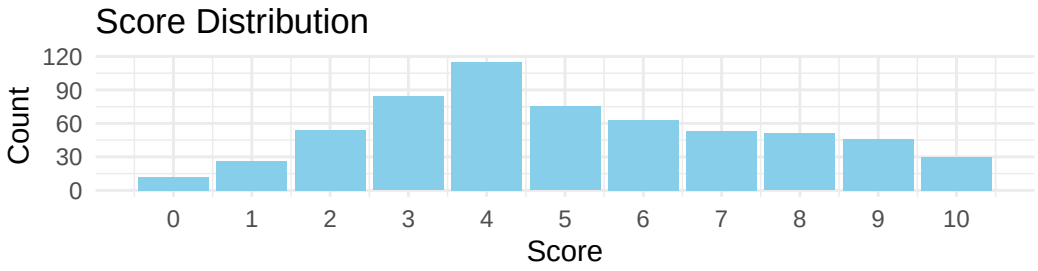


Figure 1: Distribution of total scores across dataset

As we can see in a quick summary of our data, the variable we are most interested in – the final score of users – ranges from 0 to 10, which makes sense as 10 questions can be answered either 'Yes' or 'No'. Generally, the scores are a bit right-skewed, with a single peak around 4 or 5 points. The mean score is 5.077, and the median is 5, both of which are relatively high considering that scores above 6 warrant further diagnostic evaluation. However, because suspecting a diagnosis is a reason for taking the test to begin with, these statistics are reasonable reflections of the test-taking population. We also observe that roughly 30% of test takers are encouraged to seek a diagnosis due to a score higher than 6, as calculated above. We also observe the IQR to be 4 points, as most test takers have a score between 3 and 7 points inclusive. In the context of the data, such a spread is reasonable, and no outliers exist which is understandable given the limited range of scores.

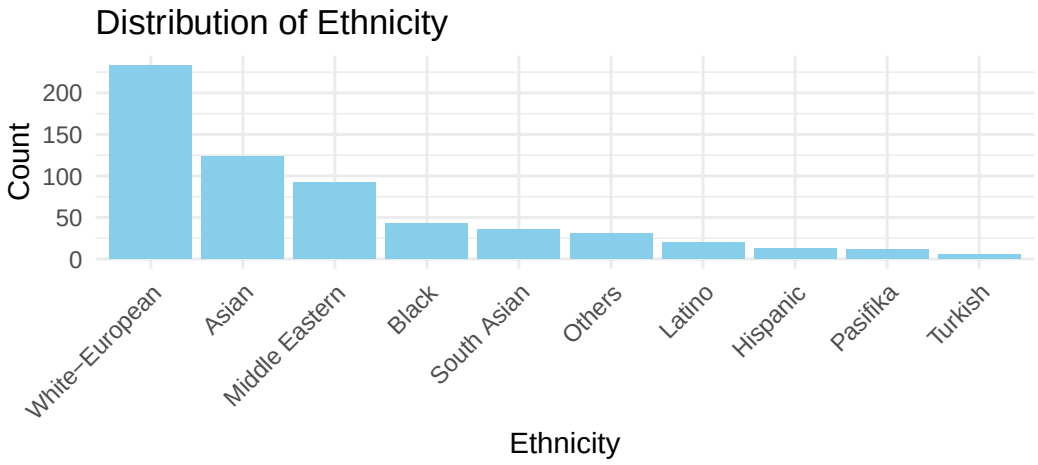


Figure 2: Distribution of ethnicities across dataset

We observe here that the most common ethnicity was White-Europeans, with over 200 observations in our dataset. Secondly and thirdly, we observe the Asian and Middle Eastern

populations to be above 100 and slightly below 100 observations, respectively. All other ethnicities have fewer than 50 observations in our data set, suggesting it may be more difficult to conclude about these populations. We can now check whether the proportion of individuals encouraged to seek a diagnosis varies by ethnicity:

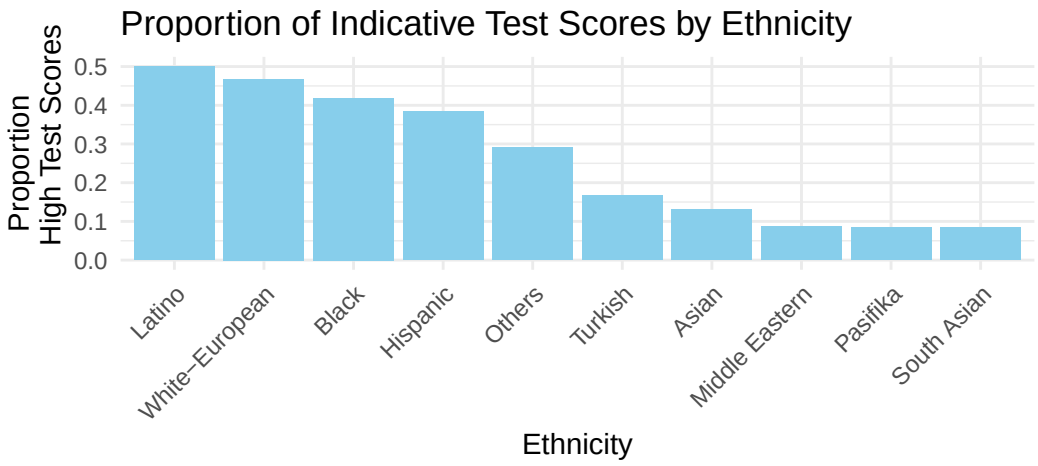


Figure 3: Proportion of indicative score across ethnicities

As we can see here, the population with the highest proportion of test takers receiving a high test score was the Latino population, with close to half achieving a score higher than a 6. Since this is a comparably uncommon population in our data set, it does raise questions of whether it is indicative of the greater Latino population or merely due to the smaller sample size in our data set. Additionally, the White-European, Black, and Hispanic populations also demonstrate a greater proportion of high test scores. The variation in these proportions according to ethnicity also warrants further exploration into how the distribution of test scores differs by ethnicity. Specifically, the questions remain of whether this means the rates of under- or over-diagnosis vary for different ethnic groups, and if so, what alterations would be necessary to ameliorate these errors.

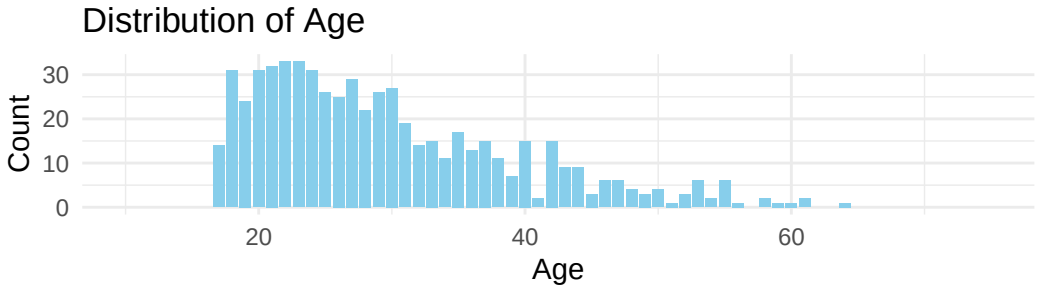


Figure 4: Age distribution across dataset

Generally, we observe the distribution of ages to possess a strong right-skewness, with a mean age of 30.22 and a median age of 27. There is a clear peak in the age distribution roughly around the early to mid-20s. The ages range from a minimum of 17 to a maximum of 383 – a false observation that should be filtered from the data. The IQR is 13 years, which is a fairly small spread given the range of ages, as we observe that the majority of test-takers are under 30 years old.

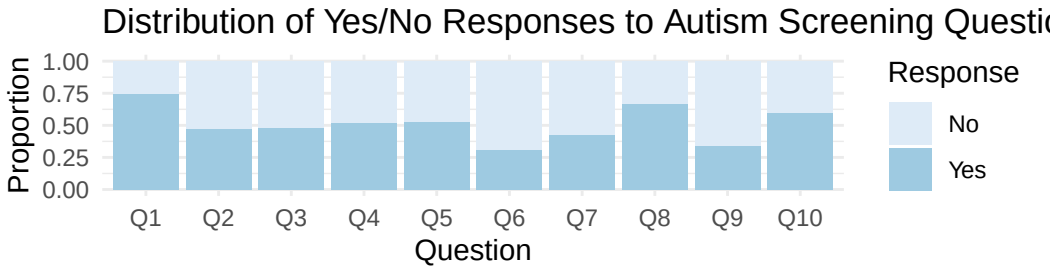


Figure 5: Distribution of Y/N responses for Q1-10. The questions are the following: Q1, noticing small sounds; Q2, finding it difficult to work out character intentions; Q3, finding it easy to read between lines; Q4, big picture-oriented; Q5, can tell if someone listening ot me is bored; Q6, can multitask; Q7, can tell feelings from faces; Q8, can go back to work when interrupted; Q9, enjoy collecting info on categories;Q10, find it difficult to work out people's intentions.

The two questions with the highest proportion of "Yes" responses were Q1 and Q8, suggesting that these behaviors are common among respondents. On the other hand, Q6 and Q9 had noticeably lower "Yes" responses, potentially indicating difficulties in such areas. Most questions, however, had roughly even proportions between "Yes" and "No" responses.

ii. Bivariate EDA

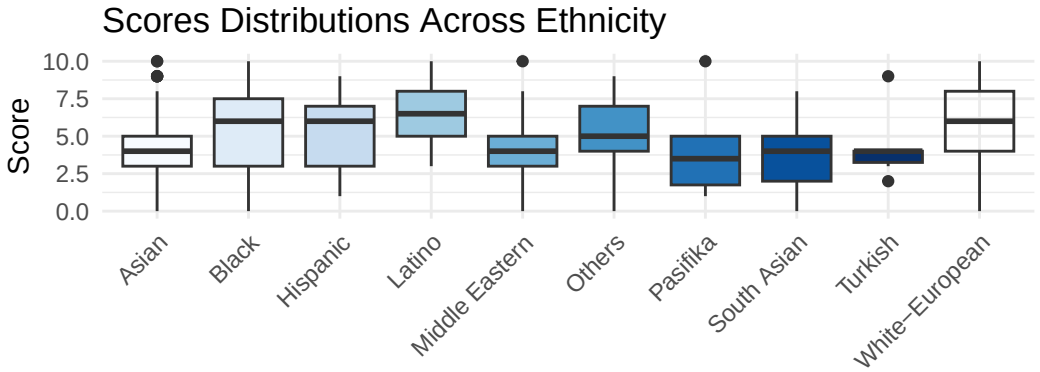


Figure 6: Score distribution across ethnicities

Here we present several graphs displaying bivariate relationships. With our first set of box plots, we find further evidence for our previous suspicion that test score distributions differ by ethnicity. Although most ethnicities have a median within the range of roughly 3 to 6, some ethnicities, like the White-European, Black, and Hispanic populations, demonstrate a greater spread through their larger IQRs, while the Turkish, Middle Eastern, and Asian populations are much more concentrated around their medians. However, most ethnicities appear to have values almost entirely across the range of 0 to 10 in their test scores.

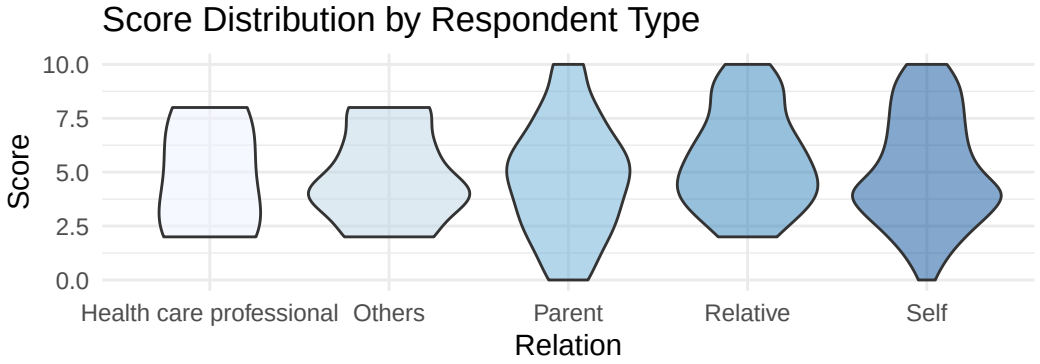


Figure 7: Distribution of scores by respondent type

Interestingly, we can also observe how the relationship between the test taker and the subject of the questions may lead to different distributions of test scores. In the case when it is filled out by a relative or health care professional for example, the observed test score is 2 or greater, while in the case of the test taker being the subject or the parent of the subject, some still received a score of 0. This could reflect how personal biases or relationships affect truthfulness during the test.

This initial exploration leads us to investigate one of these questions further, and particularly how our demographic data may impact the probability of being encouraged to seek a formal diagnosis. Though we cannot identify a definitive answer of whether it is due to social or cultural perceptions of autism within these subgroups or true difference in rates of its presence, it is worth identifying whether over- or under-diagnosis for certain subgroups is possible.

II. Methodology

i. Choosing Predictors

```
D-reduced: 739.4001
```

```
D-full: 645.2997
```

```
G-stat: 94.10049
```

```
p-value: 2.440435e-16
```

term	estimate	std.error	statistic	p.value
(Intercept)	-1.900	0.268	-7.089	0.000
ethnicityBlack	1.572	0.409	3.841	0.000
ethnicityHispanic	1.430	0.630	2.270	0.023
ethnicityLatino	1.900	0.521	3.645	0.000
ethnicityMiddle Eastern	-0.451	0.457	-0.987	0.323
ethnicityOthers	1.006	0.478	2.106	0.035
ethnicityPasifika	-0.498	1.078	-0.462	0.644
ethnicitySouth Asian	-0.498	0.660	-0.754	0.451
ethnicityTurkish	0.291	1.128	0.258	0.797
ethnicityWhite-European	1.771	0.298	5.935	0.000

```
D-reduced: 739.4001
```

```
D-fulll: 738.1941
```

```
G-stat: 1.206009
```

```
p-value: 0.8771082
```

term	estimate	std.error	statistic	p.value
(Intercept)	-1.099	1.155	-0.951	0.341
relationOthers	-0.288	1.607	-0.179	0.858
relationParent	-0.054	1.201	-0.045	0.964
relationRelative	0.351	1.224	0.287	0.774
relationSelf	0.255	1.159	0.220	0.826

```
D-reduced: 739.4001
```

```
D-full: 737.8629
```

```
G-stat: 1.537246
```

```
p-value: 0.2150282
```

term	estimate	std.error	statistic	p.value
(Intercept)	-1.059	0.190	-5.568	0.000
age	0.006	0.006	1.133	0.257

```
D-reduced: 739.4001
```

```
D-full: 734.9397
```

```
G-stat: 4.460471
```

```
p-value: 0.03468794
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.678	0.125	-5.434	0.000
genderm	-0.376	0.178	-2.108	0.035

```
D-reduced: 739.4001
```

```
D-full: 730.0734
```

```
G-stat: 9.326757
```

```
p-value: 0.002258317
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.963	0.095	-10.095	0.000
jundiceyes	0.861	0.278	3.101	0.002

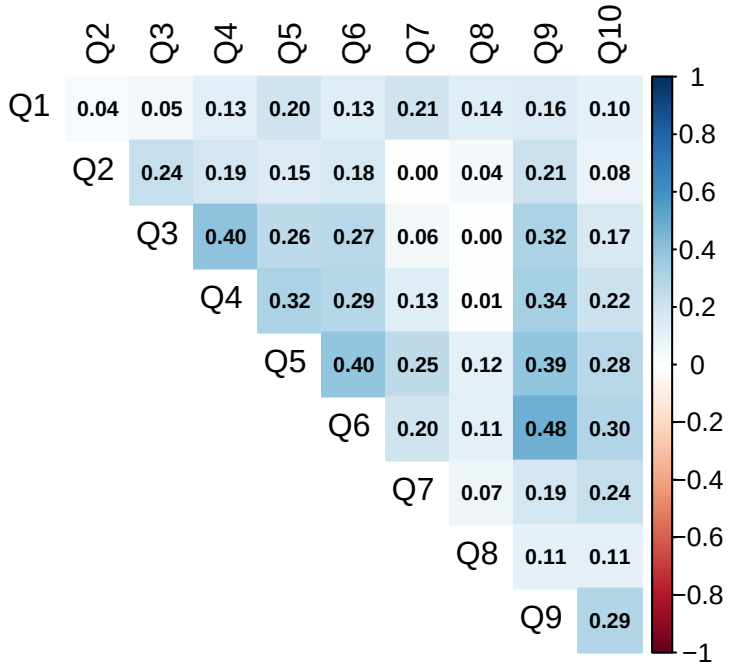
A drop-in-deviance test between a logistic null model without predictor variables and a logistic model with a single predictor was systematically conducted across ethnicity, gender, presence of neonatal jaundice, and relationship as a means of assessing which predictors provide a statistically significant improvement in model fit against the null condition. The hypothesis for this test can be observed below, where β_1 represents the coefficient for the associated predictor variable:

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

Table 1: Drop-in-deviance test results comparing null model $log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0$ to single-predictor model $log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 X_1$

From the results of this test, it is evident that the parameters that are statistically significant for model fit are ethnicity, gender, and, quite surprisingly, neonatal jaundice when compared to a null model as the p-values are either significantly less than 0.05 (as in ethnicity) or slightly below it (gender and jaundice). We further investigate the correlation between answers to questions below:



By computing a correlation heat map among the ten behavior questions to check multicollinearity, we note that surprisingly, only a few questions have stronger correlations with one another. This is rather unexpected since we would expect many behaviors to be grouped together and more correlated (e.g., social behaviors). However, the highest correlation is between questions 6 and 9, which are the questions about multitasking and liking to collect information, respectively – this is also a rather surprising pairing.

However, we note that having more uncorrelated questions is actually preferable, as that indicates that we are asking about a variety of behaviors that may lead to a diagnosis of autism, improving the chances we identify whether a subject has multiple indicators. Thus, we continue with quantifying the relationship between demographic data and high test scores.

ii. Fitting the Model

An additive logistic regression model of the following form will be fitted:

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0$$

Given that the condition we established for "high probability" (i.e. total score greater than 6) allows the logistic model to perfectly separate the response variable in the data into "high" and "low" probabilities as a binary classification, a penalized logistic regression model (LASSO) was employed to reduce the risk of overfitting (Marin, 2021).

term	estimate	std.error	statistic	p.value
(Intercept)	-0.869	0.089	-9.78	0

Then to consider adding the demographic data as potential predictors of a high score, we first fit as the following:

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_{ethnicity}X_{ethnicity} + \beta_{gender}X_{gender} + \beta_{jaundice}X_{jaundice}$$

term	estimate	std.error	statistic	p.value
(Intercept)	-1.783	0.290	-6.154	0.000
ethnicityBlack	1.542	0.411	3.748	0.000
ethnicityHispanic	1.507	0.635	2.371	0.018
ethnicityLatino	1.832	0.527	3.474	0.001
ethnicityMiddle Eastern	-0.482	0.459	-1.051	0.293
ethnicityOthers	0.998	0.480	2.079	0.038
ethnicityPasifika	-0.539	1.080	-0.499	0.618
ethnicitySouth Asian	-0.494	0.662	-0.747	0.455
ethnicityTurkish	0.361	1.130	0.320	0.749
ethnicityWhite-European	1.702	0.300	5.665	0.000
genderm	-0.291	0.195	-1.489	0.137
jundiceyes	0.640	0.305	2.099	0.036

```
D-reduced: 739.4001
```

```
D-full: 638.4442
```

```
G-stat: 100.956
```

```
p-value: 9.405191e-24
```

```
              df      AIC
model_one_fit  1 741.4001
model_two_fit 12 662.4442

              df     BIC
model_one_fit  1 745.812
model_two_fit 12 715.386
```

To compare the reduced and full models, both AIC and BIC of model two are both smaller than the values of model one. Additionally, we note that the test statistic generated by a drop-in-deviance test indicates statistical significance of adding these predictors to the model.

The intercept means that for a test taker who is an Asian woman, born without neonatal jaundice, we expect the odds of having a high score on the questionnaire to be 0.168 (calculated from exp(-1.783)).

The coefficients of ethnicity are compared to the baseline group, which is Asian. For instance, for the coefficient of `ethnicityBlack`, it means that the odds of someone identified as Black to have a high probability of autism are expected to be 4.67 (calculated from exp(1.542)) times the odds of someone identified as Asian, holding age, gender, and presence of neonatal jaundice constant.

We also note that the p-value of gender indicates a lack of statistical significance for determining a high score on the test – a surprising result given that typically, women are under-diagnosed, so we would expect there to be a lower rate of high scores. However, given that this is only a preliminary test, this could suggest that women are just as likely to seek further diagnoses as men are, but less likely to receive the necessary official diagnosis once under the care of a practitioner.

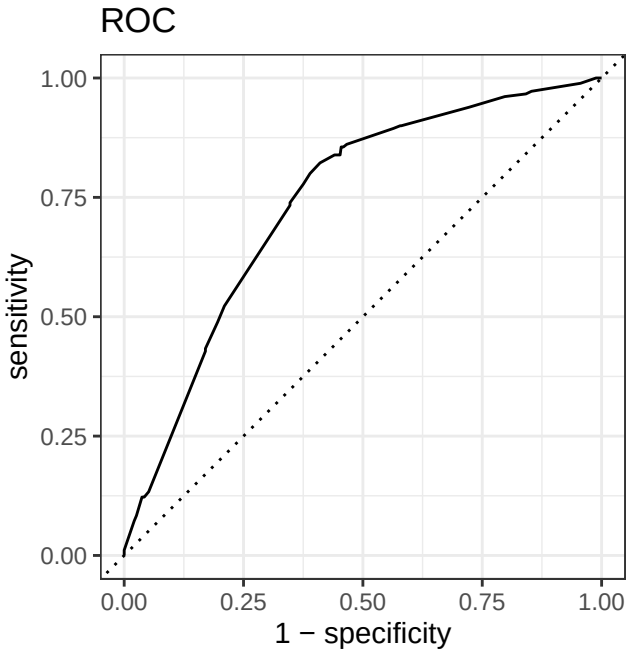
The odds of someone who was born with neonatal jaundice to have a high score on the questionnaire are expected to be 1.90 (calculated from exp(0.640)) times the odds of someone who was not born with jaundice, holding all else constant.

term	step	estimate	lambda	dev.ratio
(Intercept)	1	-1.652	0.007	0.134
ethnicityBlack	1	1.276	0.007	0.134
ethnicityHispanic	1	1.110	0.007	0.134
ethnicityLatino	1	1.517	0.007	0.134
ethnicityMiddle Eastern	1	-0.453	0.007	0.134
ethnicityOthers	1	0.681	0.007	0.134
ethnicityPasifika	1	-0.181	0.007	0.134
ethnicitySouth Asian	1	-0.341	0.007	0.134
ethnicityWhite-European	1	1.510	0.007	0.134

term	step	estimate	lambda	dev.ratio
genderm	1	-0.209	0.007	0.134
jundiceyes	1	0.543	0.007	0.134

This model was then employed to augment the original data frame and obtain the estimated probabilities for each observation, these would later be employed for an ROC curve to determine the optimal prediction threshold.

III. Results




```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.740
```

The area under the curve is 0.740, which for a model fitted to demographic data, indicates that further investigation into how these questions may over- or under-diagnose certain populations could still be worth pursuing. However, as that is still a fairly low AUC, it could also indicate that this particular questionnaire does a relatively good job of avoiding under- or over-diagnosis, and could include questions that are a strong reference for future screening.

For model assumptions, logistic regression requires the assumption of log-odds possessing linearity, randomness, and independence of observations. It is worth noting that of these conditions, randomness is unlikely to be satisfied due to the nature of the data set – it was collected by people who already were likely to suspect an ASD diagnosis, as they are those who took the test voluntarily. However, our other two assumptions are likely to hold for this model, and we still determine our questions, model, and results worth further investigation. We do not need to check linearity as none of our predictors are quantitative and our all categorical, we do not need to. We can also assume independence as there are no anticipated temporal or spatial relationships between individuals taking the questionnaire.

Once more, a .740 AUC for factors independent of autistic behaviors indicate that further research is required in the diagnostic criteria and formulation of questionnaires of for autism – but could also suggest that perhaps this particular test is less biased compared to others. It is worth ensuring that everyone can receive the appropriate diagnosis, and thus, receive any desired accommodations for improvement of quality of life.

IV. References

- Aishworiya, R., Kim, V., MA, Stewart, S., Hagerman, R., & Feldman, H. M. (2023). Meta-analysis of the Modified Checklist for Autism in Toddlers, Revised/Follow-up for Screening. *PEDIATRICS*, *151*(6). https://doi.org/10.1542/peds.2022-059393
- Curnow, E., Utley, I., Rutherford, M., Johnston, L., & Maciver, D. (2023). Diagnostic assessment of autism in adults – current considerations in neurodevelopmentally informed professional learning with reference to ADOS-2. *Frontiers in Psychiatry*, *14*. https://doi.org/10.3389/fpsyt.2023.1258204
- Hirota, T., & King, B. H. (2023). Autism spectrum Disorder. *JAMA*, *329*(2), 157. https://doi.org/10.1001/jama.2022.23661
- Marin, Z. (2021, April 26). GLM fit: Algorithm did not converge – How to fix it. Statology. https://www.statology.org/glm-fit-algorithm-did-not-converge/

::: Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML. :::

V. Additonal Materials

i. Raw Model Data