A Model on Intelligence Scores

STATS 101A Summer20

Team 11: Sylvia Ma, Albert Cao, Lia Bergman - Turnbull, Jack Barrera, Mingxuan Gao, Helen Zhang

Abstract

A person's intelligence level is a product of a series of complicated and multifaceted genetic and social factors. Many researchers have been able to draw a connection between one's intelligence level and socioeconomic status as well as the level of education of the person and his family. Our research attempts to analyze the factors that affect a stroke patient's intelligence by constructing a multiple linear model and a logistic regression model with a wider array of factors, such as BMI, level of cholesterol, and race. Our models show that intelligence level is highly influenced and determined by a person's level of education received and race. Although BMI and high cholesterol are also found to be statistically significant variables, they do not explain much about our model because of the interaction effect between the two variables. Socioeconomic status, on the other hand, can be statistically significant when examined alone with the intelligence score but is less influential in our model. Our logistic model also confirms our finding that race and levels of education are important predictors in estimating a person's intelligence level. Further improvements for our research include expanding the dataset to the entire population and reducing the interaction effect between high cholesterol and BMI.

Question to be answered:

Many research have shown that both innate and acquired intelligence determines a person's overall intelligence level. Although around 40% to 80% of a person's intelligence is inherited, the idea that environmental factors also determine intelligence levels has long been established.[1] In a research published in 1987 by Reynold et al at Texas A&M University, race,

[1] Sternberg, R. J. (2012). Book review. *Intelligence, 40*(5), 509-510. doi:10.1016/j.intell.2012.05.001

education, and occupation were found to be highly correlated to a person's intelligence level.[2] The sample that Reynold et al was based on WASI-R, which is an updated version of WASI, the intelligence variable we used in our model. Also, whether people's health risks are related to their intelligence is another interesting question being asked over time. Therefore, our analysis attempts to investigate how health and demographic factors such as race, education, socioeconomic status, BMI, and high cholesterol affect the prediction of a person's intelligence score by analyzing in multiple linear and logistic regression models.

Outcome and Predictors

Our dataset is named stroke, which has 474 observations and 25 variables. Our outcome variable is a numeric variable named Wechsler Abbreviated Scale of Intelligence score(WASI). It is an individual intelligence test widely used and approved in psychological research. In our model, we use the percentile rank of WASI as the indicator of a person's intelligence.

To predict WASI, we chose four predictors, of which two are numeric and two are categorical. The numeric variables are BMI and Education since several studies have investigated the prospective associations between intelligence and various social factors, such as educational attainment, occupation, income, and BMI.[3] Specifically, BMI is a calculated parameter that characterizes a person's body fat, and can be used to determine whether a person is obese or overweight. The higher the BMI, the higher the body fat composition. Education is another numeric variable that is a self-report of the highest grade that a person has completed. Education goes from 9 to 24, which corresponds to 1st year high school and doctoral study.

With respect to the categorical variables, we choose the binary high cholesterol variable, the binary race variable, and socioeconomic status score (SES) with three levels. Previous research has shown that cholesterol is essential for brain maturation[4] and is often related to high BMI, so we choose high cholesterol as a predictor, denoted a 0 to those who do not have high cholesterol and a 1 to those who have high cholesterol. Also, since race is an important factor related to the controversy over the causes of intelligence--genetic or environmental, we use race as a binary predictor, having "African American" as baseline and "White" as the second level. In addition, we transformed a numeric variable, socioeconomic status score(SES), into a factor with 3 levels. The SES is influenced by income, education, and occupation status, which ranges from

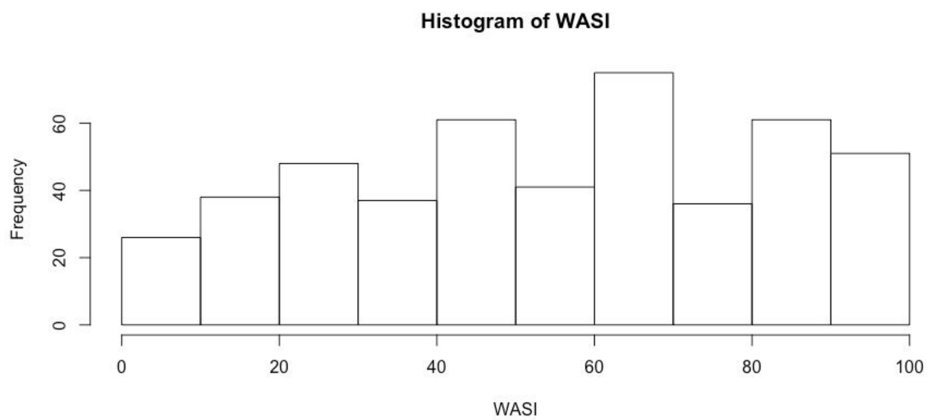
---

[2] Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & Mclean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology, 25*(4), 323-342. doi:10.1016/0022-4405(87)90035-5
[3] Halkjær, J., Holst, C., & Sørensen, T. I. (2003). Intelligence test score and educational level in relation to BMI changes and obesity. *Obesity Research*, 11(10), 1238-1245.
[4] Saher, G., Brügger, B., Lappe-Siefke, C., Möbius, W., Tozawa, R. I., Wehr, M. C., ... & Nave, K. A. (2005). High cholesterol level is essential for myelin membrane growth. *Nature neuroscience*, 8(4), 468-475.

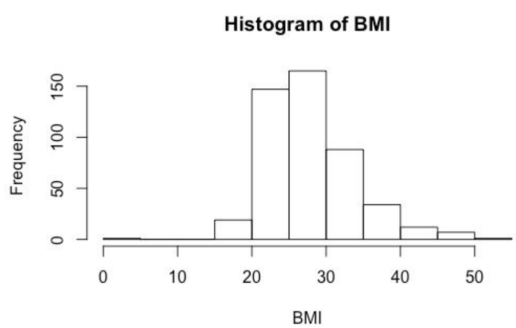
14 and 75 in our data. Because of the skewness of SES, we transformed it into a categorical variable.

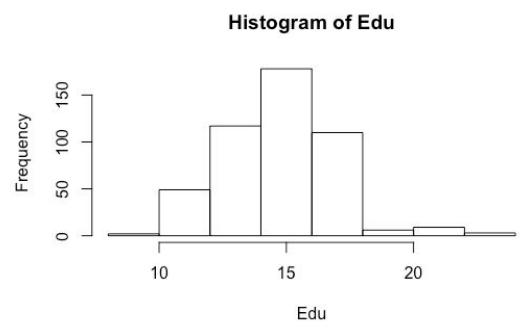
Exploratory Data Analysis

Our dependent variable is the Wechsler Abbreviated Scale of Intelligence score(WASI). Its distribution looks approximately normal, so we don’t make any transformation.

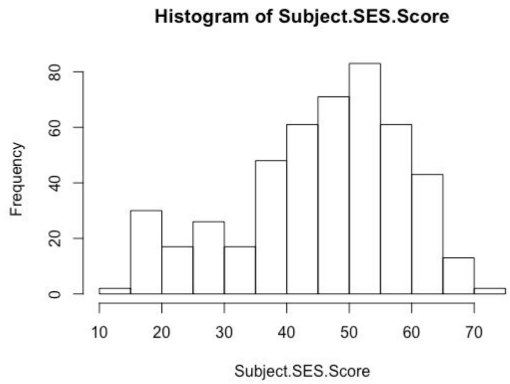


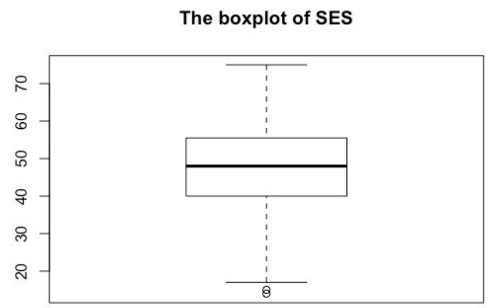
Education, and BMI are the two numeric variables we chose to include in our model. The histogram of these two variables show that they closely resemble a normal distribution.





As shown below, the distribution of socioeconomic status score(SES) is negatively skewed, so we transform it into a factor with 3 levels. We denote SES ranging from 14-42 as “Low”, 42-52 as “Medium”, and 52-75 as “High”.

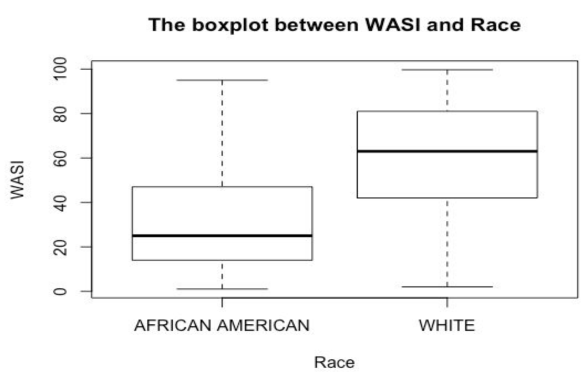


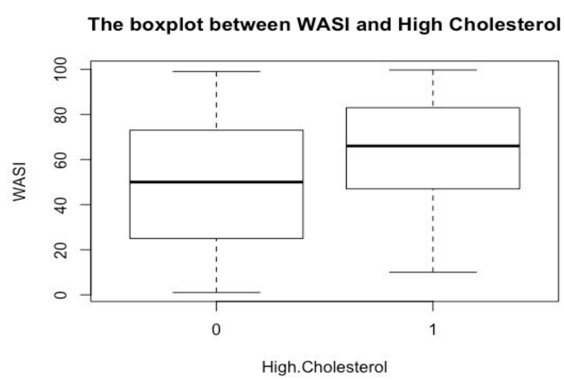


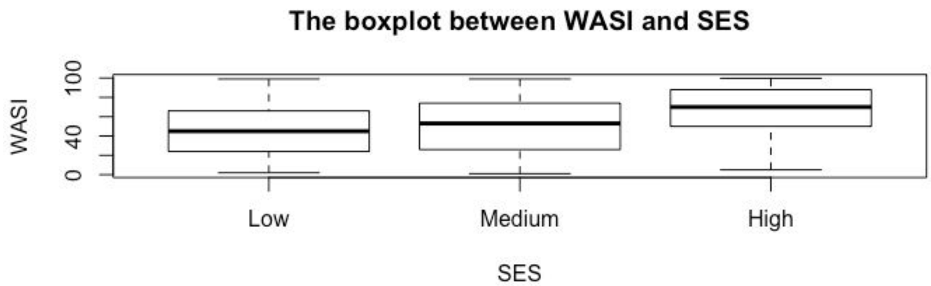
The following table shows the number of observations in each categorical variables:
Table I. Summary of Categorical Variables

Variables	Baseline - # of Observations	The 2nd level - # of Observations	The 3rd level - # of Observations
Race	African American - 82	White - 392	
High Cholesterol	Not High Cholesterol - 323 (code as 0)	High Cholesterol - 151 (code as 1)	
SES	Low - 152	Medium - 148	High - 173

According to the table, we confirm that the frequency of each cell is large enough in a sample size of 474. The boxplots below show that all of our categorical variables have noticeable differences in mean of WASI depending on different levels.







Correlation Matrix

Our correlation matrix, which has a r value of -0.025, shows that there exists a minimal degree of multicollinearity between the two numeric variables we include in our model: Education and BMI.

Table II. Correlation Matrix

	Edu	BMI
Edu	1	-0.025
BMI	-0.025	1

However, by checking the Variance Inflation Factor(VIF) for all predictors, we find that both high cholesterol and its interaction with BMI are higher than 5. It is due to the significant interaction effect between BMI and high cholesterol in our model. Since it helps us to better understand the combined effect of health and social factors on intelligence, we still keep them in our model.

Table III. VIF

Race	Education	SES	BMI	High Cholesterol	BMI:High Cholesterol
1.07	1.26	1.15	1.20	5.04	5.12

Final Linear Regression Model

Our final regression model is lm(WASI~Race+Edu+SES+High.Cholesterol*BMI).

Table IV. Summary of Multiple Linear Regression

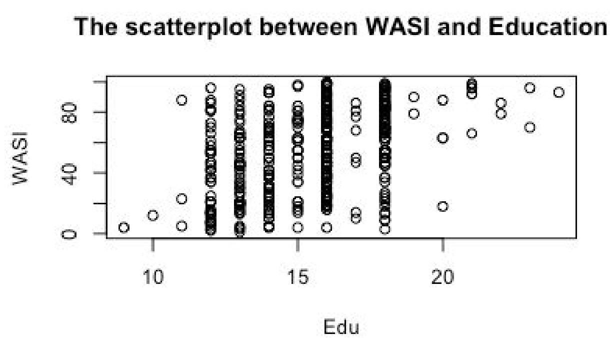
Predictors	Coefficient	Standard error	t-value	p-value
Constant	-4.1255	10.3608	-0.40	0.6907
RaceWHITE	21.0093	2.9677	7.08	5.4e-12 ***
Education	3.7408	0.5823	6.42	3.3e-10 ***
SESMedium	-5.6651	2.8984	-1.95	0.0512 .
SESHigh	0.0939	3.3631	0.03	0.9777
BMI	-0.6186	0.2133	-2.90	0.0039 **
High Cholesterol (1=have)	-18.8853	11.3891	-1.66	0.0980 .
BMI:High Cholesterol	0.9274	0.3916	2.37	0.0183 *

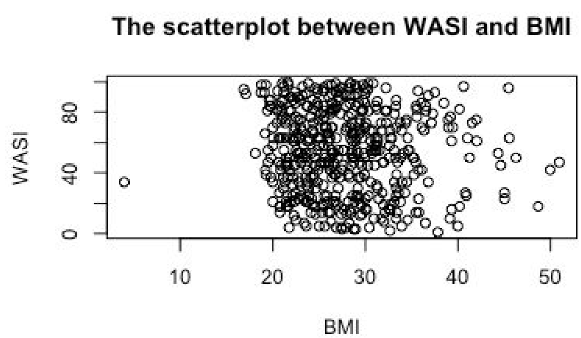
.p< .1 *p< .05 **p< .01 ***p< .001

F-statistic: 27.5 on 7 and 465 DF

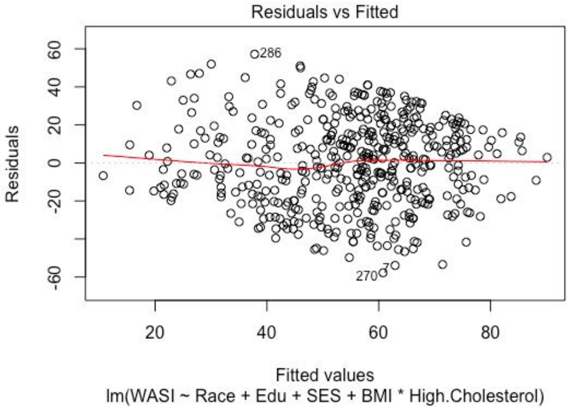
Multiple R-squared: 0.293,   Adjusted R-squared: 0.282

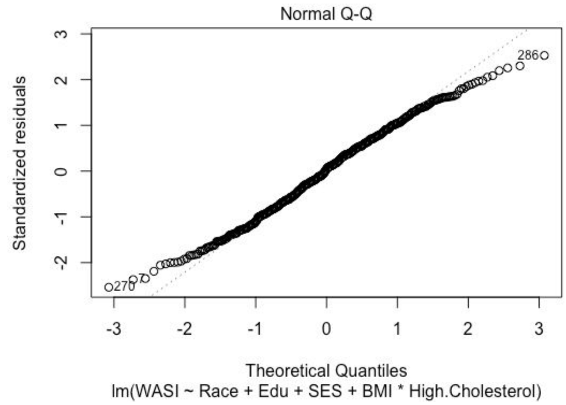
Test of the assumptions





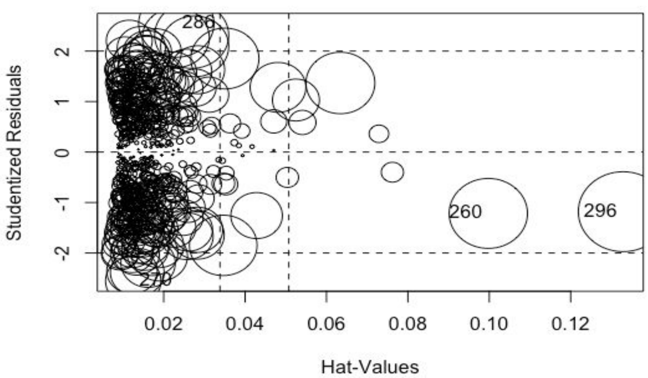
The scatterplot indicates a linear pattern in the data, so we meet the assumption of linearity in the multiple linear regression.





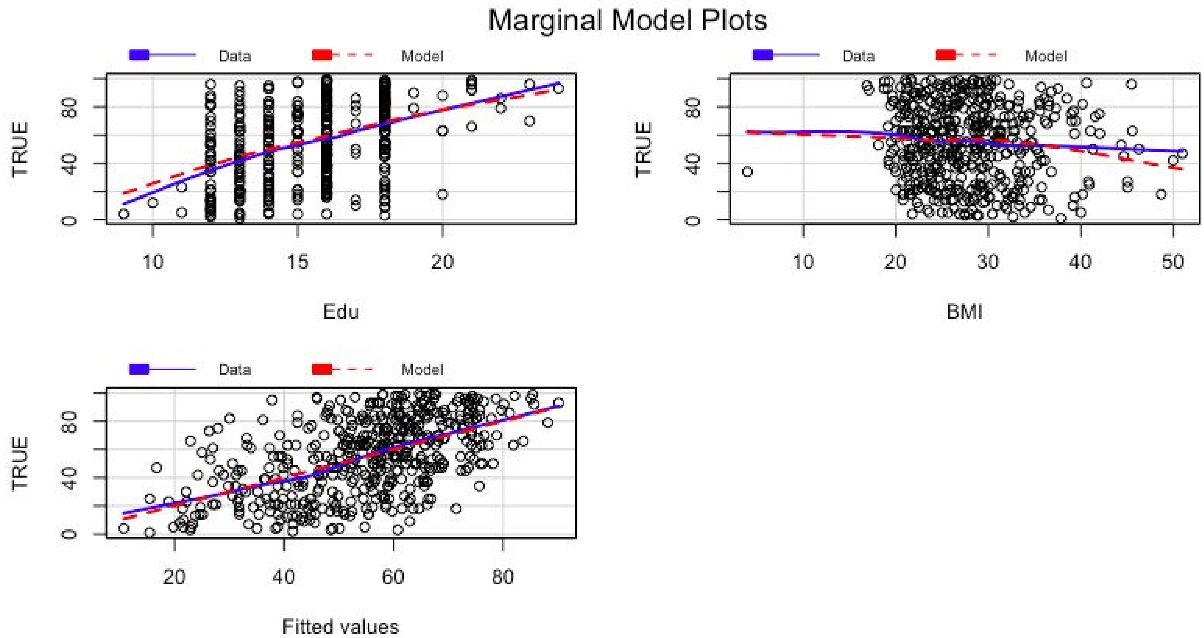
The Residuals vs. Fitted plot does not have any noticeable pattern or any fan shape. It confirms the linearity in our sample and also suggests homoscedasticity. By doing the Non-constant Variance Score Test(NCV Test), we get a p-value of 0.2 (0.2>0.05) and fail to reject the null hypothesis that the variance is constant in our sample. Therefore, we meet the assumption of variance constancy.

In the QQ plot, most of the points are on the line indicating that the quantiles from the sample are close to quantiles from the normal model. Therefore, we also meet the assumption of normality.



	Standardized residual	Hat	CookD
260	-1.22	0.0998	0.0205
270	-2.55	0.0125	0.0102
286	2.54	0.0231	0.0189
296	-1.18	0.1329	0.0267

The influential plot shows that there are four potential influential points in our sample. Since our sample size is large, we consider +-4 as the cutoff of the outlier. Therefore, none of the leverage points are bad leverage points here and there is no need to remove them.



In all of the three Marginal Model Plots(MMP plots) above, all of the nonparametric estimates given as solid blue curves are almost indistinguishable from the smooth fitted values shown as dashed red lines. Thus, we conclude our model is valid for WASI.

Model Selection

Table V. AIC and R square adjusted for each subset model

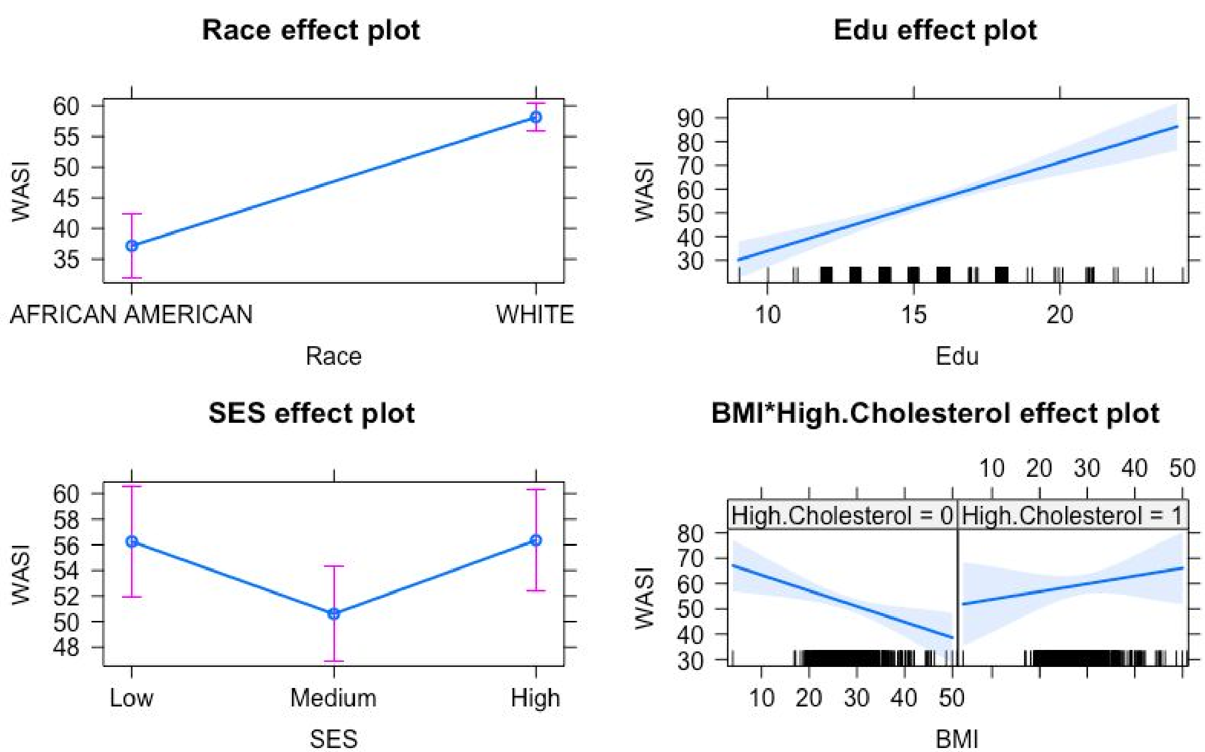
Model	**Predictors**	**AIC**	**R2-adjusted**
M1	SES	4412	0.107
M2	SES+Education	4374	0.179
M3	SES+Education+Race	4326	0.259
M4	SES+Education+Race+BMI	4325	0.262
M5	SES+Education+Race+BMI+High Cholesterol	4318	0.275
M6	SES+Education+Race+BMI*High Cholesterol	4314	0.282

The best model is M6 as it has the lowest AIC and highest adjusted R-squared. The highest adjusted R-squared shows that our final model includes useful predictors to explain the most of the variance of WASI in our sample. The lowest AIC indicates that our model is not overfitting.

Interpretation of Results

Table VI. R square change provided by each predictors

Race	Education	SES	BMI	High Cholesterol	BMI:High Cholesterol
14.5%	11.0%	1.0%	0.39%	1.5%	0.85%

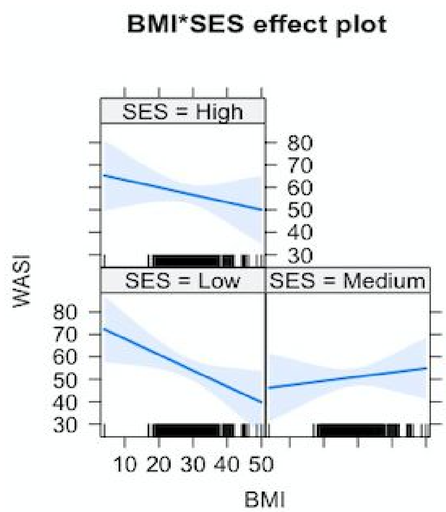


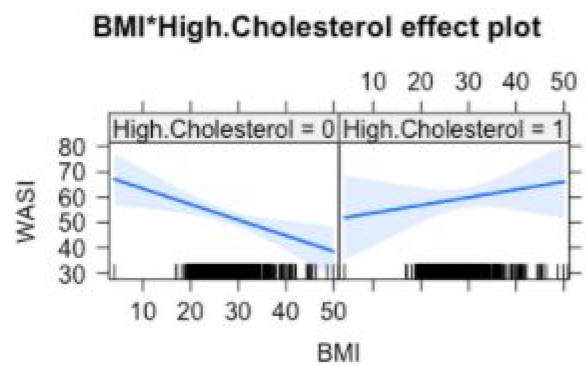
Keeping all else constant…

The white participants, on average, scored 21.01 points more on intelligence compared to African American participants. This finding is statistically significant. Based on many previous researches, social factors contribute to one's intelligence. So, we believe that the overall low SES and education level among African Americans might negatively impact their intelligence score. Since their mean duration of education is 1.3 years lower and the mean SES is 10 points lower than the white participants, these social factors could negatively influence their intelligence score.

For every additional year of education, on average, one’s intelligence score is predicted to be 3.74 points higher. This finding is statistically significant and intuitive. There are two ways to interpret it. We might say that people who are able to achieve higher degrees usually have higher intelligence scores. Alternatively, higher levels of education might help people to achieve higher scores in the intelligence test. It is not certain which one is the reason and which one is the result.

For every point increase in BMI, on average, one’s intelligence score is predicted to be 0.62 points lower. This finding is statistically significant. To further examine BMI, we constructed an interaction effect plot between BMI and high cholesterol. It is surprising to find that the effect of BMI on a person’s intelligence score depends on whether or not the person suffers from high cholesterol. For those who do not have high cholesterol, their intelligence scores decrease while their BMI increases. Conversely, for those who have high cholesterol, their intelligence scores increase while their BMI increases. Although the overall BMI is negatively correlated to intelligence score, the impact of BMI varies based on the predictor high cholesterol.





Since the SES of people who suffer from high cholesterol is 8.7 points higher than those who don’t, we further investigate whether the interaction effect between BMI and high cholesterol is similar to that of BMI and SES. It turns out that for those who have low SES, their intelligence scores also decrease while their BMI increases. It shows that increasing BMI has a stronger negative correlation with intelligence for low income people. However, whether high SES people are more likely to suffer from high cholesterol is still controversial. Some articles claims that persons higher in the socioeconomic hierarchy command greater access to resources to avoid s cardiovascular disease risk factors,[5] but others show that high calorie diets are more common among high SES.[6] This is a topic that needs further research.

According to the coefficient in the final model, on average, one's intelligence score is predicted to be 18.89 points lower if the person suffers from high cholesterol. Conversely, in the simple linear regression between high cholesterol and WASI, the intelligence score is predicted to be higher for those suffering from high cholesterol. This change in the sign of coefficient happens after adding the interaction effect between BMI and high cholesterol(from M5 to M6). Since the combined effect has a stronger impact on intelligence score than the main effect of each predictors, the effect of high cholesterol itself on WASI becomes almost insignificant. In general, we conclude that high cholesterol plays a complex role in determining one's intelligence. In a paper by Bernard G. Schreurs, the author also recognizes the complexity and mixed effect of cholesterol on intelligence and acknowledges the contrasting conclusions that have been made in the field.[7]

Table VII. Unstandardized Regression Coefficients Predicting WASI

Predictors	M1	M2	M5	M6
SESMedium	5.86*	-1.00	-5.57.	-5.67.
SESHigh	20.91***	7.43*	-0.01	0.09
Education		4.03***	3.79***	3.74***
RaceWHITE			20.30***	21.01***
BMI			-0.34.	-0.62**
High Cholesterol (1=have)			7.48**	-18.89.
BMI:High.Cholesterol				0.93*
R square	0.11	0.18	0.28	0.29
R square change		0.07	0.1	0.01

[5] Winkleby, M. A., Cubbin, C., Ahn, D. K., & Kraemer, H. C. (1999). Pathways by which SES and ethnicity influence cardiovascular disease risk factors. *Annals of the New York Academy of Sciences*, 896(1), 191-209.

[6] Rosero-Bixby, L., & Dow, W. H. (2009). Surprising SES gradients in mortality, health, and biomarkers in a Latin American population of adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* 64(1), 105-117.

[7] Schreurs, B. G. (2010). The effects of cholesterol on learning and memory. *Neuroscience & Biobehavioral Reviews,34*(8), 1366-1379. doi:10.1016/j.neubiorev.2010.04.010

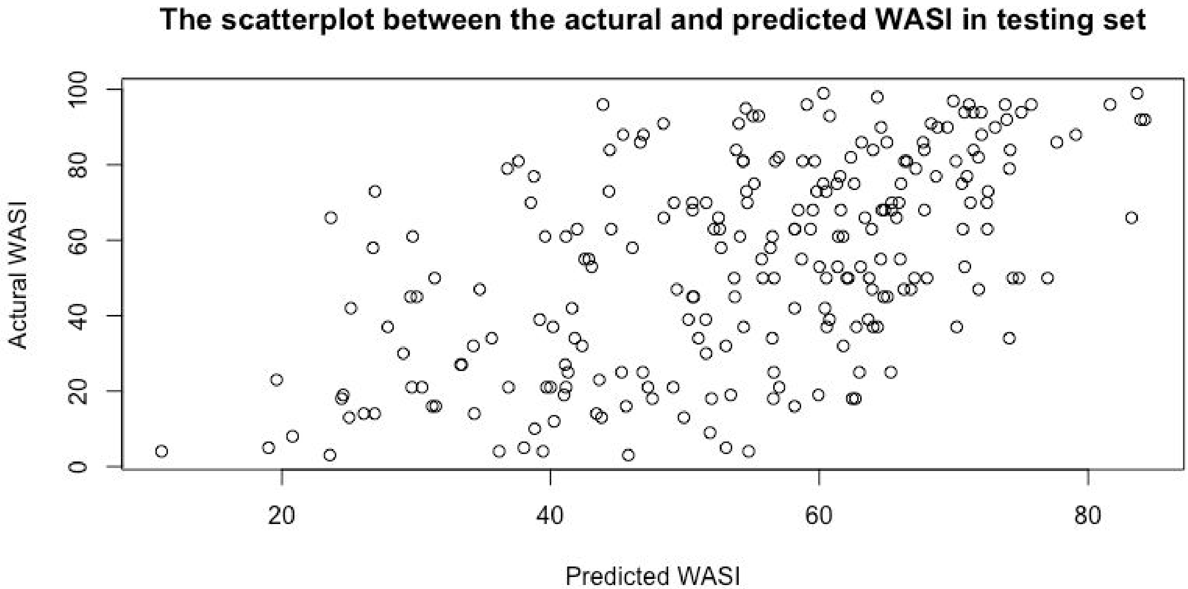
.p< .1 *p< .05 **p< .01 ***p< .001

Socioeconomic status is not a statistically significant predictor in our model. However, when being examined alone with the outcome, SES is a good predictor which can explain about 11% of variance in the intelligence score(WASI). In the simple linear regression between SES and WASI, the intelligence score is predicted to be higher when one's social status increases. Our finding verifies the conclusion made in a study by Tabriz et al, where the researchers also concluded that IQ is highly correlated to the income level.[8] However, as we add education in our model, the effect of SES on WASI becomes insignificant and the direction of their correlation changes(from M1 to M2). According to the coefficient in our final model, switching from people who have low SES to medium SES, on average, one's intelligence score is predicted to be 5.67 points lower, while switching from people who have low SES to high SES, one's intelligence score is predicted to be 0.09 points higher. This is largely due to the multicollinearity between SES and education in which SES can explain about 37% of variance in education. In fact, education is usually used as a contributor to SES in many articles. Since education has a stronger correlation with the intelligence score than SES, the coefficient of SES becomes misleading in this model.

Cross Validation

In order to evaluate how our model performs in prediction, we apply the cross validation to see the extent to which the prediction model created based on the training sample is valid for the testing model and similar data sets. We equally divide our data into two parts as the 'training sample' and the 'testing sample'. The coefficients found from the model developed on the training sample was used to predict WASI is the testing sample. The coefficient of correlation between actual WASI and predicted WASI in the testing model is around 0.475. This is a moderate correlation. However, our result is acceptable given the fact that a person's intelligence is not solely dependent upon race and social factors. There are still many variables, such as the intelligence of one's parents, that are not included in our model.

[8] Tabriz, A. A., Sohrabi, M., Parsay, S., Abadi, A., Kiapour, N., Aliyari, M., . . . Roodaki, A. (2015). Relation of intelligence quotient and body mass index in preschool children: A community-based cross-sectional study. *Nutrition & Diabetes, 5*(8). doi:10.1038/nutd.2015.27



Logistic Regression Model

For our logistic regression model, we used the same outcome variable of WASI score and we chose two numerical predictors: Education and BMI, and two categorical predictors: Race and High Cholesterol. For the outcome variable, above median WASI score was coded as 1 and below median WASI score was coded as 0 in order to create a binary outcome variable. The output of the model and the coefficient estimates follow below:

Table VIII. Summary of Logistic Regression

Predictors	Coefficient	Standard error	t-value	p-value
Constant	-0.512435	0.178641	-2.869	0.00431**
RaceWHITE	0.297653	0.057853	5.145	3.94e-07***
Education	0.055314	0.009520	5.810	1.15e-08***
BMI	-0.003555	0.003561	-0.998	0.31869
High Cholesterol	0.103429	0.046643	2.217	0.02707*

*p< .05 **p< .01 ***p< .001

Table IX. Confidence Interval

Predictors	Estimate	2.5%	97.5%
Constant	0.5990	0.4221	0.8502
RaceWHITE	1.3467	1.2023	1.5084
Education	1.0569	1.0373	1.0768
BMI	0.9965	0.9895	1.0034
High Cholesterol	1.1090	1.0121	1.2151

Interpretation of Results:

Keeping all else constant, the odds of having an above median WASI score is 34.67% higher for those who are white vs any other race.

Keeping all else constant, for each year of increase in education, the odds of having an above median WASI score is increased by 5.69%.

Keeping all else constant, for each point of increase in BMI, the odds of having an above median WASI score is increased by 1.35%.

Keeping all else constant, the odds of having an above median WASI score is 10.9% higher for those who have high cholesterol.

Comparison of the two models

Overall, our multiple linear regression model and logistic model draw similar conclusions. In both models, race, education level, and whether or not a person has high cholesterol are all strong and statistically significant predictors.

One exception is that BMI functions well as a predictor under the multiple linear regression model but becomes statistically insignificant under the logistic model. Still, overall the two models convey similar conclusions about the relationship between these predictors and WASI score.

Possible Improvements

Since the problem of multicollinearity exists between SES and Education, we couldn't interpret how SES itself impacts intelligence prediction. If we could adjust SES with education so that it only contains information about one's income and occupation instead of education level, we could better investigate how each social factor makes a change in one's intelligence.

Also, if we could have more innate predictors such as one's parents' education level and income, we could have a more accurate model about one's intelligence and further analyze how much percent of intelligence is contributed by innate factors.

Additionally, if we could have a dataset that is not limited to stroke patients, we might be able to expand our conclusion to a bigger population.

Conclusion

The multiple linear regression model and the logistic model that we created demonstrate that race, education, BMI, and high cholesterol are statistically significant predictors of one's intelligence score. Socioeconomic status, however, can be a statistically significant predictor when examined alone. One area for improvement in our study is the interpretation of high cholesterol. The contrasting results it creates lead to additional questions on the role that cholesterol plays on one's intelligence. Also, the fact that socioeconomic status is correlated to other predictors, such as education, reduces its significance and is a shortcoming of our model.

Bibliography

HALKJÆR, JYTTE, CLAUS HOLST, AND THORKILD I.A. SØRENSEN. Intelligence test score and educational level in relation to BMI changes and obesity. *Obes Res.* 2003;11:1238-1245.

Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & Mclean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology, 25*(4), 323-342. doi:10.1016/0022-4405(87)90035-5

Saher, G., Brügger, B., Lappe-Siefke, C., Möbius, W., Tozawa, R. I., Wehr, M. C., ... & Nave, K. A. (2005). High cholesterol level is essential for myelin membrane growth. *Nature neuroscience*, 8(4), 468-475.

Schreurs, B. G. (2010). The effects of cholesterol on learning and memory. *Neuroscience & Biobehavioral Reviews,34*(8), 1366-1379. doi:10.1016/j.neubiorev.2010.04.010

Sternberg, R. J. (2012). Book review. *Intelligence, 40*(5), 509-510. doi:10.1016/j.intell.2012.05.001

Tabriz, A. A., Sohrabi, M., Parsay, S., Abadi, A., Kiapour, N., Aliyari, M., . . . Roodaki, A. (2015). Relation of intelligence quotient and body mass index in preschool children: A community-based cross-sectional study. *Nutrition & Diabetes, 5*(8). doi:10.1038/nutd.2015.27

Rosero-Bixby, L., & Dow, W. H. (2009). Surprising SES gradients in mortality, health, and biomarkers in a Latin American population of adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 64(1), 105-117.

Winkleby, M. A., Cubbin, C., Ahn, D. K., & Kraemer, H. C. (1999). Pathways by which SES and ethnicity influence cardiovascular disease risk factors. *Annals of the New York Academy of Sciences,* 896(1), 191-209.