

CHAPTER FOUR: PRESENTATION AND DISCUSSION OF RESULTS

Contents

CHAPTER FOUR: PRESENTATION AND DISCUSSION OF RESULTS	1
4.1 Introduction	1
4.2 Data Structure and Case Processing	2
4.2 Hypothesis Testing	2
4.2 Block 0 Analysis	3
4.3 Regression Analysis and Hypotheses Testing	4
4.4 Correlation and Intercorrelation Analysis	5
4.5 Goodness of Fit Test Statistics	5
4.5.1 OMNIBUS Test of Model Coefficient	6
4.5.2 Hosmer and Lemeshow Test	6
4.6 The Contingency Table for Hosmer and Lemeshow Test	6
4.7 Model Summary	7
4.8 Classification Table	7
4.9 Conclusion	8
References	8

4.1 Introduction

In the prediction of Dementia (Alzheimer's disease in particular), a variety of factors have been investigated over the past decade in the UK and across the globe. In line with the research questions guiding this study, experimental data was collected from Oasis (2022). This data was examined to determine which factors have the greatest impact on Alzheimer's disease prediction as well as the cumulative predictive power of the adopted empirical model. According to the NHS England (NHS, 2021), dementia occurs when there is a decline in the brain function of an individual. Alzheimer's disease which is the most common type of Dementia is caused by the abnormal build-up of two types of proteins namely amyloid and tau (Nixon & Yang, 2011). In the broad Dementia literature, research evidence indicates that genetic factors, advanced age, and several environmental factors are responsible for AD

expression (Munoz & Feldman, 2000). Therefore, in line with the research questions guiding this study, and the associated hypotheses, logistic regression was adopted to analyse various predictive factors of AD. A logistic regression indicates the criterion variable which uses a dummy variable indicating two major predictive groups. SPSS version 28 has been used for this analysis.

4.2 Data Structure and Case Processing

A total of 373 cases have been involved in this study. Out of this number 371 representing, 99.5% were used while there were 2 missing cases representing 0.5%. The dependent variable which is AD is measured as a dichotomous variable which is binary measured as Dementia=1, No Dementia=0. These are the two broad prediction groups. These participants were at one point diagnosed with AD under a series of observations which led to their diagnoses. One categorical variable which is male (M), and Female (F) was identified and treated as such. While a total of 161 males were involved, the total number of females are 210. The analysis, therefore, represents group membership regarding AD. Table 4.1 below presents the case processing summary for the study.

Table 4.1: Case Processing Summary

Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	371	99.5
	Missing Cases	2	.5
	Total	373	100.0
Unselected Cases		0	.0
Total		373	100.0

4.2 Hypothesis Testing

In line with the research questions, as indicated in the research methodology, the following hypotheses are designed to be tested through logistics regression. A total of ten (10) hypotheses have been tested. Table 4.2 below indicates the various hypotheses guiding this study.

Table 4.2: Case Processing Summary

Null Hypothesis	Description
H_1	MR Delay has a significant positive impact on the development of Alzheimer's disease

H_2	M/F has a significant positive impact on the development of Alzheimer's disease
H_3	Age has a significant positive impact on the development of Alzheimer's disease
H_4	EDUC has a significant positive impact on the development of Alzheimer's disease
H_5	SES has a significant positive impact on the development of Alzheimer's disease
H_6	MMSE has a significant positive impact on the development of Alzheimer's disease
H_7	Clinical Dementia Rating (CDR) has a significant positive impact on the development of Alzheimer's disease
H_8	Estimated Total Intracranial Volume (eTIV) has a significant positive impact on the development of Alzheimer's disease
H_9	Normalized whole brain volume (nWBV) has a significant positive impact on the development of Alzheimer's disease
H^{10}	Atlas scaling factor (ASF) has a significant positive impact on the development of Alzheimer's disease

4.2 Block 0 Analysis

In understanding a logistic regression, two blocks of analysis are usually conducted regarding the predictors and the outcome variable. These are Block 0 and Block 1. Block 0 indicates the results of the analysis without any of the independent variables in the model. This is called the baseline data analysis. There are no predictors engaged in this stage of analysis. In Block 0, analysis, all the predictors are excluded from the model to study the model. As much as this model is not the most important model of interest, it indicates the significance of the predictors outside the model. In block 0 whiles M/F, Educational level, MMSE, CDR and nWBV are statistically significant at 1%, MR delay is statistically significant at 5%. Table 2 below shows the block 0 analysis. The results also indicate that Age, SES and ASF are statistically insignificant.

Table 2: Block 0 variable analysis

			Score	df	Sig.
Step 0	Variables	MR Delay	5.619	1	.018
		M/F (1)	18.372	1	.000
		Age	.010	1	.920
		EDUC	13.258	1	.000
		SES	1.445	1	.229
		MMSE	102.169	1	.000
		CDR	224.191	1	.000
		eTIV	.403	1	.525
		nWBV	36.451	1	.000

		ASF	.154	1	.695
--	--	-----	------	---	------

4.3 Regression Analysis and Hypotheses Testing

This is the block of importance. All predictor variables are included in the analysis. It shows how each of the predictors affects the outcome variable. In this analysis, all ten variables have been examined. Various statistics of these variables include the coefficient (*B*), Standard Error, Wald, Degrees of freedom, significance, and the odds ratio. Table 4.3 below shows the regression analysis of the predictive model.

Table 4.3: Block 1 variable analysis

Variable		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1 ^a	MR Delay	.002	.001	7.709	1	.005	1.002	1.001	1.003
	M/F(1)	-1.259	.780	2.609	1	.106	.284	.062	1.308
	Age	-.005	.046	.011	1	.917	.995	.909	1.090
	EDUC	.220	.144	2.349	1	.125	1.247	.940	1.652
	SES	1.145	.370	9.575	1	.002	3.142	1.522	6.490
	MMSE	.218	.231	.897	1	.343	1.244	.792	1.955
	CDR	17.141	2.938	34.046	1	.000	.000	.000	.000
	eTIV	.032	.014	5.267	1	.022	1.033	1.005	1.061
	nWBV	8.621	11.169	.596	1	.440	5547.291	.000	178.280
	ASF	31.331	16.684	3.527	1	.060	404.836	.255	6429.000
Constant	-100.735	42.021	5.747	1	.017	.000			
a. Variable(s) entered on step 1: MR Delay, M/F, Age, EDUC, SES, MMSE, CDR, eTIV, nWBV , ASF .									

In Block 1 analysis, while MR delay, SES and CDR are statistically significant at a 1% level, eTIV is significant at a 5% level. In addition, ASF is partially significant at the 10% level. The other variables namely M/F, Age, MMSE, and nWBV are statistically insignificant. This

implies that a unit increase in MR delay, SES and CDR increases in AD development by 002%,1.145% and 17.141%. Subsequently, a unit increase in eTIV and ASF increase dementia development by .032% and 31.331% respectively. Thus, the null hypotheses H_1 H_5 H_7 are accepted. Again, H_8 and H_9 are accepted, while H_2 , H_3 H_4 H_6 and H_{10} are unsupported or rejected.

4.4 Correlation and Intercorrelation Analysis

Correlation analysis indicates how the variables of interest are related to each other. Usually, correlations exist between 0 and 1/-1 where 0.1- 0.3 represents a weak correlation, 0.3-0.5 represents a medium correlation 0.6-0.9 represents a strong correlation. A correlation of +1 or -1 represents a perfect positive or negative correlation respectively. A correlation of 0 indicates that there is no correlation which signifies weak multicollinearity in the data. Table 4.4 below shows the correlation matrix of the variables.

Table 4.4: Correlation Matrix

Correlation Matrix												
		Constant	MR Delay	M/F(1)	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Step 1	Constant	1.000	-.139	.161	-.178	-.003	-.146	-.240	.436	-.943	-.214	-.927
	MR Delay	-.139	1.000	-.044	-.075	.051	.074	.031	-.469	.137	.048	.121
	M/F(1)	.161	-.044	1.000	-.002	-.201	-.313	.080	.183	-.225	.000	-.114
	Age	-.178	-.075	-.002	1.000	.159	.036	.142	.083	-.060	.719	-.103
	EDUC	-.003	.051	-.201	.159	1.000	.616	-.076	-.144	-.087	.169	-.106
	SES	-.146	.074	-.313	.036	.616	1.000	-.013	-.360	.136	-.062	.092
	MMSE	-.240	.031	.080	.142	-.076	-.013	1.000	.197	.035	.197	.034
	CDR	.436	-.469	.183	.083	-.144	-.360	.197	1.000	-.521	.034	-.449
	eTIV	-.943	.137	-.225	-.060	-.087	.136	.035	-.521	1.000	-.056	.982
	nWBV	-.214	.048	.000	.719	.169	-.062	.197	.034	-.056	1.000	-.113
	ASF	-.927	.121	-.114	-.103	-.106	.092	.034	-.449	.982	-.113	1.000

4.5 Goodness of Fit Test Statistics

The goodness of fit measure determines whether the model adequately describes the data in the study. There are two main goodness of fit tests which are conducted. These are the OMNIBUS Test of Model Coefficient and the Hosmer and Lemeshow Test. These tests are presented below.

4.5.1 OMNIBUS Test of Model Coefficient

The purpose of the OMNIBUS Test of Model Coefficient is to test whether the model is significant compared to block 0. If the model is significant, it shows that there is a significant improvement in fit compared to the null model (Block 0). Since all the two blocks are statistically significant at a 1% level, the predictive model of this study shows a good fit and can be relied upon for other predictive studies. Table 4.5 below shows the OMNIBUS Test of Model Coefficient

Table 4.5: Omnibus Tests of Model Coefficients

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	407.292	10	.000
	Block	407.292	10	.000
	Model	407.292	10	.000

4.5.2 Hosmer and Lemeshow Test

Hosmer and Lemeshow Test is also another test of model fit which is employed in a logistic regression analysis. The Hosmer and Lemeshow Test indicate a poor model fit if its significance is less than 0.05. In this study, the model significance is 0.966 which is higher than the 0.05 threshold. Table 4.6 below shows the results of the Hosmer and Lemeshow Test

Table 4.5:The Hosmer and Lemeshow Test

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	2.409	8	.966

4.6 The Contingency Table for Hosmer and Lemeshow Test

In analysing the contingency table, there should be no difference between the observed and predicted results. As indicated in table 4.6 below, the model is confirmed as reliable and valid due to the observed values truly reflecting the predicted values.

Table 4.6: Contingency Table

Contingency Table for Hosmer and Lemeshow Test						
		Dementia = 1.00		Dementia = 2.00		Total
		Observed	Expected	Observed	Expected	
Step 1	1	37	37.000	0	.000	37
	2	37	36.994	0	.006	37
	3	37	36.960	0	.040	37
	4	37	36.802	0	.198	37
	5	24	22.678	13	14.322	37
	6	6	6.114	31	30.886	37
	7	1	2.659	36	34.341	37
	8	1	1.259	36	35.741	37
	9	1	.445	36	36.555	37
	10	0	.090	38	37.910	38

4.7 Model Summary

The model summary indicates the Psuedo-R Square to explain the variance in the dependent variable attributable to the predictor variables. Two main analyses are usually done in understanding the total variance explained by the predictor variables. These are the Cox& Snell R^2 which shows the percentage of change attributable to the predictors. The Negekerke's R^2 also known as the adjusted represents the true value of change after the model adjustment. In table 4.7 below, the model explains the variance in the outcome variable by 66.6% and its adjusted value is 88.9%. Thus the total predictive power of the model is 66.8% before adjustment.

Table 4.7: Model Summary

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	106.805 ^a	.666	.889

4.8 Classification Table

Finally, the classification table indicates how well the model can predict the correct category of the investigation by the predictors. This is usually compared to block 0. This classification also shows the accuracy at which each category is predicted as well as the number predicted. As observed in Table 4.8 below, the module can predict the dementia group at a 90.1% rate of

success (163 participants) accurately and the non-dementia group at 98.9% accurately. The combined prediction rate is therefore at 94.6%.

Table 4.7: Classification Table of Dementia Prediction

Classification Table ^a					
	Observed		Predicted		
			Dementia		Percentage Correct
			1.00	2.00	
Step 1	Dementia	1.00	163	18	90.1
		2.00	2	188	98.9
	Overall Percentage				94.6

a. The cut value is .500

4.9 Conclusion

The analysis conducted in this chapter depicts the use of a logistic regression analysis which is used in estimating dichotomous variables in a study where the outcome variable is binary. The study aimed at investigating ten (10) main predictive factors of AD. The logistic regression, therefore, estimated the probability of participants falling into dementia and non-dementia groups by observing the odds ratio. The Block 1 analysis indicates that while the null hypotheses H_1 H_5 H_7 H_8 and H_9 are accepted, H_2 , H_3 H_4 H_6 and H_{10} are unsupported or rejected. More importantly, the module can predict the dementia group at a 90.1% rate of success (163 participants) accurately and the non-dementia group at 98.9% accurately. The combined prediction rate is therefore at 94.6%.

References

- Munoz, D. G., & Feldman, H. (2000). Causes of Alzheimer's disease. *Canadian Medical Association journal*, 162(1), pp. 65 - 72.
- NHS. (2021, January 11). *Causes of dementia*. Retrieved from NHS: <https://www.nhs.uk/conditions/dementia/causes/>
- Nixon, R. A., & Yang, D.-S. (2011). Autophagy failure in Alzheimer's disease—locating the primary defect. *Neurobiology of disease*, 2011, Volume 43, Issue 1, 43(1), pp. 38 - 45.

DO NOT COPY