Linear Regression Analysis on NIH Stoke Scale

Xin Xie[1], Xingjian Xie[1], Yijin Zhao[1], Yue Wang[2]

[1]Department of Biostatistics, Fielding School of Public Health,
[2]Department of Epidemiology, Fielding School of Public Health.

Introduction

- Stroke, a type of cardiovascular disease, is one of the leading causes of death in the United States.
- NIH Stroke Scale, as a measurement, subjectively quantifies the severity of the stroke and provides guidance for treatment.

Objective

We would like to investigate whether there is an association between NIH Stroke Scale at 90 Days and 3 months Lesion Volume among stroke patients, after controlling for potential confounders.

Methods

Participants:

- The data we use is the National Institute of Neurologic Disease and Stroke's Archived Clinical Research data. There are 100 separate variables recorded for 624 stroke patients enrolled in the NINDS study. Patients in the NINDS study were randomized to treatment and placebo groups. The treatment was rt-PA.
- Based on literature review and common knowledge, we chose 11 variables. 177 patients were dropped due to missing values and outliers, resulting in 447 patients in the final model.

Statistical analysis:

- The outcome of interest is 90 days NIH Stroke Scale, and the predictor of interest is 3 months Lesion Volume.
- Due to positive skewed distribution, we use log transformation on 90 days NIH Stroke Scale and 3 months Lesion Volume to make them normally distributed.
- NIH stroke Scale at 90 days was centered by adding 1 before log transformation to remove 0.
- Simple linear regressions were conducted on each predictor and outcome, as well as a full model with all variables. Coefficients, adjusted and unadjusted and R^2 are reported.
- R Package 'corrplot' was used to draw a correlation map.
- VIF for 13 variables in full model are calculated to diagnose multicollinearity.
- Race and gender have greater P-values, but we still include them in the model based on background knowledge.
- The 2 best subsets of predict variables were chosen using smallest Cp, AIC, BIC values in SAS 9.4.

Results

- White people accounts 61.7% of total participants, the average age of the participants is 67.
- The average length of stay is 13 days.
- The distribution of male and female are relatively equal.
- Result of the best subset contains 8 variables: Lesion volume, Gender, Age, NIH stroke scale at 24 hours. Length of stay, Hypertension, and NIH stroke at 7-10 days, with R2 equals 0.6626 with the smallest CP, AIC and BIC.
- Best subset 2 keeps the variable treatment. The reason is that there might be relationship between treatment and the 90 days NIH stroke scale based on background knowledge.
- Black people tends to have higher mean NIH stroke scale at 90 days controlling for other variables.
- Gender and Treatment of Medicine are not very significant in this model.

Table 1: Sample Characteristics

Variable	Count(percentage)	Mean	Range	SD
log(NIH Stroke Scale at 90 Days*)	447(100%)	1.9	(0, 3.7)	1.2
NIH stroke scale at 90 Days*	447(100%)	13.2	(1, 42)	14.2
log(3 months Lesion Volume)	447(100%)	3.2	(-2.2, 6.4)	1.9
3 months Lesioin Volume	447(100%)	67.5	(0.1, 594)	89.8
NIH stroke scale at baseline	447(100%)	15.4	(2, 37)	6.9
Age (year)	447(100%)	67.4	(26.5, 89)	11.4
NIH Stroke Scale at 24 Hours	447(100%)	13.2	(0, 37)	8.2
Length of stay (day)	447(100%)	13.2	(1, 88)	11.5
NIH Stroke Scale at 7-10 Days	447(100%)	11.9	(0, 40)	9.4
Race				
Black=1	130(29.08%)	—	—	—
White=2	276(61.71%)	—	—	—
Hispanic=3	31(6.94%)	—	—	—
Asian=4	5(1.12%)	—	—	—
Other=5	5(1.12%)	—	—	—
Gender				
Male=0	250(55.9%)	—	—	—
Female=1	197(44.1%)	—	—	—
Treatment group				
Received placebo=0	234(52.35%)	—	—	—
Received rt-PA=1	213(47.65%)	—	—	—
Hypertension				
No hypertension	143(31.99%)	—	—	—
With hypertension	304(68.01%)	—	—	—

* NIH Stroke Scale at 90 Days is added by 1 to remove 0

Table 2
Linear Regrssion Result on NIH Stroke Scale at 90 days
Bivarate and Multivariate Results, Including Best Subsets Models

	Bivariate (n=447)	Multivariate (n=447)	Best Subset 1 (n=447)	Best Subset 2 (n=447)
	Parmeter Estamite Beta (se), p-value	Parmeter Estimate Beta (se), p-value	Parmeter Estimate (se), p-value	Parmeter Estimate (se), p-value
log(3 months Lesion Volume[1])	0.309(0.026), <0.0001	0.085(0.024), 0.0004	0.085(0.023), <0.001	0.085(0.023), 0.0003
Gender	-0.011(0.117), 0.927	-0.116(0.070), 0.099	-0.118(0.069), 0.088	-0.116(0.069), 0.094
NIH stroke scale at baseline	.086(0.007), <0.0001	0.004(0.007), 0.562	—	—
Age (year)	0.025(0.005), <0.0001	0.009(0.003), 0.008	0.008(0.003), 0.013	0.008(0.003), 0.010
NIH Stroke Scale at 24 Hours	0.106(0.005), <0.0001	0.016(0.008), 0.053	-0.019(0.008), 0.013	0.018(0.008), 0.019
Treatment	-0.177(0.116), 0.126	-0.084(0.070), 0.230	—	-0.079(0.069), 0.252
Length of stay (day)	0.038(0.005), <0.0001	0.008(0.003), 0.017	0.008(0.003), 0.008	0.008(0.003), 0.009
Hypertension	0.306(0.124), 0.014	0.177(0.076), 0.020	0.182(0.075), 0.0157	0.182(0.075), 0.016
NIH Stroke Scale at 7-10 Days	0.102(0.004), <0.0001	0.073(0.007), <0.0001	0.073(0.007), <0.0001	0.0730.007), <0.0001
Race				
Black=1	0.1830.127), 0.152	0.177(0.079), 0.0266	0.164(0.077), 0.034	0.160(0.077), 0.039
Hispanic=3	0.190(0.228), 0.406	0.070(0.138), 0.613	—	—
Asian=4	0.556(0.551), 0.313	0.344(0.331), 0.299	—	—
Other=5	0.413(0.551), 0.455	0.239(0.331), 0.472	—	—
R^2	—	0.665	0.663	0.664
Adjusted R^2	—	0.655	0.656	0.657
CP	—	—	7.543	8.233
AIC	—	—	-287.037	-286.380
BIC	—	—	-284.607	-283.841

[1]NIH Stroke Scale at 90 Days is added by 1 to remove 0.

Conclusion

- This study investigated the association between NIH Stroke Scale at 90 Days and 3 months Lesion Volume while controlling for potential confounders using data from the NINDS dataset. The results showed that 3 months Lesion Volume is a strong predictor of 90 days NIH Stroke Scale.
- The results showed that black people tend to have a higher mean NIH Stroke Scale at 90 days.
- There is a significant positive relationship between 3 months lesion volume, age, 24-hour stroke scale, Length of Stay, Hypertension, Race of Black and NIH Stoke Scale at 90 days.
- Gender was not found to be significant in this model.
- Overall, this study provides important insights into the relationship between NIH Stroke Scale and lesion volume in stroke patients, but further research is needed to fully understand the complexities of this relationship.

Limitations

- The dataset only includes patients in the NINDS study and met certain inclusion criteria, which may not reflect the broader stroke population.
- The study only includes patients who agreed to participate in the study, which may introduce self-selection bias.
- The study only examines the association between NIH Stroke Scale and 3 months Lesion Volume, and it is unclear whether these findings can be extrapolated to other outcomes or time points.
- The NIH Stroke scale relies on the judgement of the investigators, which may be subject to misclassification bias.
- The model includes some variables with high correlation, but we keep them in model since their VIF values are smaller than 10.