0.
0.3886898

MATH4432 Mini-Project 1 : Regression: Animal Species Sleeping Hours

Chow Wing Ho(20279607)

This project is going to explore what may affect the sleep that animal needs which using the multiple regression to find out the result. Also, this project use cross-validation and bootstrap to the estimation of test error.

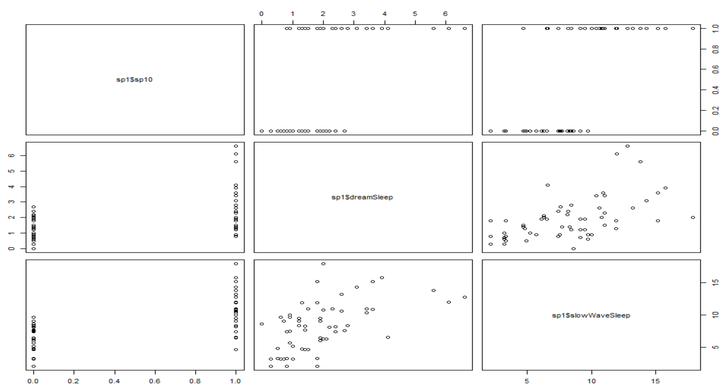
Methods

- Missing values

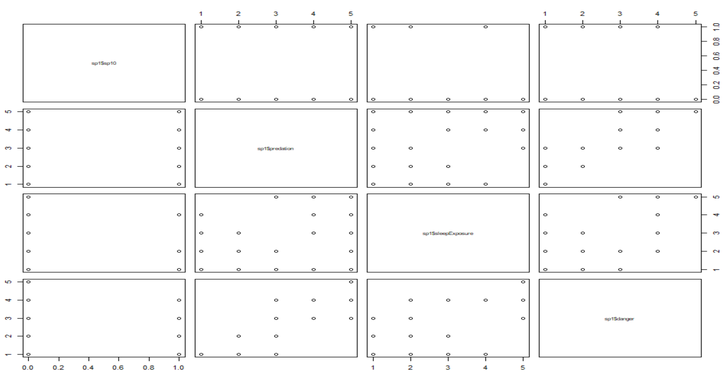
- using the packages “Mice” ,max. iteration and CART decision tree to fill in missing value which this method use the missing value be Y and other values be X to predict the value Y.

- Dependent Variable
- Use dummy variable (sp10) to express the dependent variable which the value of sleep larger than median value of sleep
- Independent Variables
- Use the multiple regression of all factors to decide the predictors with p-values for the dependent variable.
- equation:
- m9 = lm(sp10 ~ slowWaveSleep + dreamSleep + brain + body + life + gestation + predation + danger + sleepExposure)
- Summary of m9 regression show that the slowWaveSleep and dreamSleep are significant to Dependent Variable and the variables (danger, predation and sleepExposure) are likely having relationship to the Dependent Variable.

Sp10 vs slowWaveSleep and dreamSleep



Sp10 vs danger , predation and sleep Exposure



Test the multilinear regression

- Use slowWaveSleep, dreamSleep , predation ,danger and sleepExposure for different various multilinear regression (M1 – M10)

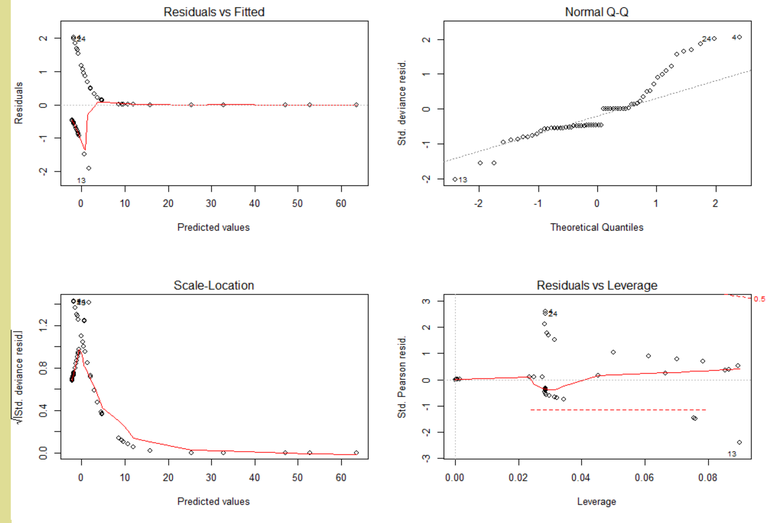
- M1 = glm(sp10 ~ dreamSleep + slowWaveSleep + predation + danger + sleepExposure)
- M2 = glm(sp10 ~ slowWaveSleep + danger+ predation + dreamSleep + sleepExposure^2)
- M3 = glm(sp10 ~ slowWaveSleep + danger^2+ predation + dreamSleep + sleepExposure)
- M4 = glm(sp10 ~ slowWaveSleep + danger + predation^2+ dreamSleep)
- M5 = glm(sp10 ~ slowWaveSleep + dreamSleep +danger^2 * predation^2)
- M6 = glm(sp10 ~ slowWaveSleep^2 + dreamSleep + danger + predation)
- M7 = glm(sp10 ~ slowWaveSleep + dreamSleep^2 + danger + predation)
- M8 = glm(sp10 ~ slowwavesleep^2 * dreamSleep^2 + danger + predation)
- M9 = glm(sp10 ~slowWaveSleep +dreamSleep)
- M10= glm(sp10 ~ slowwavesleep^2 * dreamSleep^2)

Results of M1 –M10

	Dependent variable: Sleep (sp10)									
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
slowWaveSleep^2						0.055***				
						(0.02)				
dreamSleep	1.942**	1.569**	1.876**	1.198**	1.137**	1.181**			1.262**	
	(0.91)	(0.73)	(0.86)	(0.58)	(0.57)	(0.58)			(0.55)	
sleepExposure^2		0.16								
		(0.12)								
predation*danger					(0.00)					
					(0.00)					
slowWaveSleep	1.091***	1.008***	1.068***	0.828***	0.818***		0.818***		0.817***	
	(0.36)	(0.34)	(0.35)	(0.28)	(0.26)		(0.27)		(0.26)	
danger^2			-0.385*							
			(0.22)							
dreamSleep^2							0.279*			
							(0.16)			
slowwavesleep*dreamSleep								0.008***		0.009***
								(0.003)		(0.003)
predation	0.69	0.47	0.83			0.14	0.16	0.78		
	(0.92)	(0.87)	(0.80)			(0.77)	(0.73)	(0.63)		
danger	-1.69	-1.28		-0.62		-0.42	-0.47	-0.98		
	(1.19)	(1.12)		(0.73)		(0.83)	(0.82)	(0.71)		
sleepExposure	1.20		1.36							
	(0.72)		(0.72)							
predation^2				0.07						
				(0.11)						
Constant	-12.656***	-10.188***	-14.385***	-8.086***	-8.572***	-5.369***	-7.149***	-1.779*	-8.955***	-2.220***
	(4.43)	(3.40)	(4.85)	(2.80)	(2.61)	(2.02)	(2.43)	(1.08)	(2.57)	(0.57)
Observations	62	62	62	62	62	62	62	62	62	62
Note:	*p<0.1; **p<0.05; ***p<0.01									

The table show that variable related to danger, predation and sleep exposure are not significant which p-value larger than 0.01,0.05,0.1.

The diagram of best regression (M10)



Cross-validation and bootstrap

- After testing the different multilinear regression, use the “boot” packages with cv.glm () to compare the MSE with ten regression
  - cv.glm (sp1,Mi)$Delta # “i” is integer 1 - 10
- The result finds that the Mean Squared Error of the tenth multilinear regression is the lowest (0.3886898)
- Use bootstrap method to test error of model (M10)
- The result of bootstrap that standard deviation of M10 is quite low which is 0.003579715

Conclusion

- In conclusion:
- 1) the simple multilinear regression of nine variables can show some variables are useless in affecting sleep of animals (e.g.. Brain, Body…) as the coefficient of these useless variables are extremely small.
- 2) from the result of ten regression, we can see that the p-values of variables relate to slowWaveSleep and dreamSleep are mostly very small and significant in P<0.01, that means these two variables are key factor to affect the sleep of animals
- 4) the standard deviation of regression M10 by using bootstrap is very small that means fill in the missing value of NAs are appropriate.