Practical 7: Logistic Regression Model Diagnostics (Binomial Data)

Data were collected on a random sample of people to ascertain what factors influenced behavioural change during a Swine Flu outbreak. The variables collected were:

Behaviour   Whether the person changed behaviour relating to hygiene (Yes or No).

Age   Age category of the person.

Severe   Whether the person believed the consequences of contracting Swine Flu were severe.

Information   Whether the person believed the information available about Swine Flu was adequate.

The aggregated data are as follows:

Severe	Information	Age	n	Changed Behaviour
Yes	Adequate	18-40	18	11
		40-50	23	14
		50-60	22	11
		60+	17	5
	Not Adequate	18-40	19	4
		40-50	35	15
		50-60	30	8
		60+	22	8
No	Adequate	18-40	24	10
		40-50	37	13
		50-60	29	8
		60+	24	6
	Not Adequate	18-40	28	11
		40-50	42	14
		50-60	37	15
		60+	30	9

Read the data into R using 5 separate vectors, applying appropriate numeric codes for categorical variables as follows:

```
Severe <- c(1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2)
Info <- c(1,1,1,1,2,2,2,2,1,1,1,1,2,2,2,2)
Age <- c(1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4)
n <- c(18,23,22,17,19,35,30,22,24,37,29,24,28,42,37,30)
CB <- c(11,14,11,5,4,15,8,8,10,13,8,6,11,14,15,9)
```

Fit an additive logistic regression model, with Severe, Information and Age as explanatory variables.

1. What is the sample size?

2. What is the deviance of the fitted model?

Calculate the residuals and other diagnostic statistics for the model.

3. What is the value of the leverage for the first case?

4. What is the value of Cook's Distance leverage for the first case?

5. What is the value of the Pearson Residual for the first case?

6. What is the value of the Deviance Residual for the first case?

7. Compare the values of the Pearson and Deviance Residuals for all cases. What do you conclude?

Obtain an index plot of the deviance residuals versus case number.

8. Using the index plot of the deviance residuals, what do you conclude about the systematic component of the model?

9. Using the Pearson Residuals, how many outliers are there?

10. Using the Deviance Residuals, how many outliers are there?

11. What is the cut-off value for Leverage?

12. How many cases of high leverage are there?

13. How many cases of high influence are there?

14. Which case has the highest influence?

15. Why has this case high influence?

16. Which single case is most worthy of investigation?

17. Is there a need to investigate any other case?