# Tutorial 8 On-site Questions

(Logistic Regression)

Data set `Smarket.csv` contains data on percentage returns for the S&P 500 stock index over 1250 days, from the beginning of 2001 until the end of 2005. For each date, the data contains the trading volume and the percentage returns for each of the five previous trading days, Lag1 through Lag5.

1. Find the proportion of the days that the direction is up.
2. Fit a logistic regression model (called model M3) that helps to predict the direction of S&P 500 index, based on trading volume and the percentage returns for each of the five previous trading days, Lag1 through Lag5.
3. Consider $\delta = 0.5184$ as the border line to predict the direction be up or down, calculate the accuracy of model M3 when predicting for the days in the data set given.

   *Hint: To recall on what the role of $\delta$ is, please revise Tutorial 5 question 2*
4. (*Extra*) Write R code to plot the ROC curve for model M3. Derive and report AUC value of it.