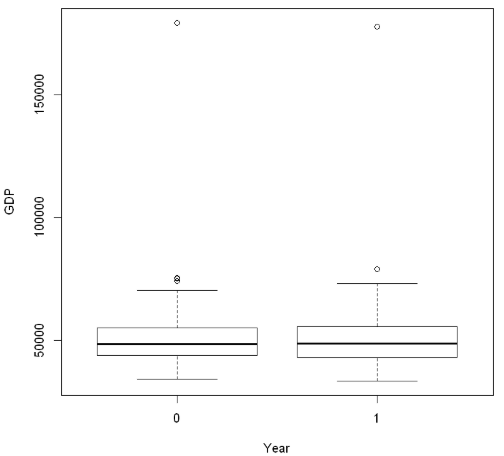
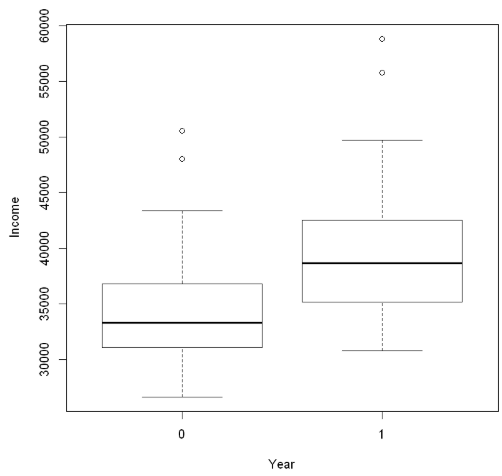
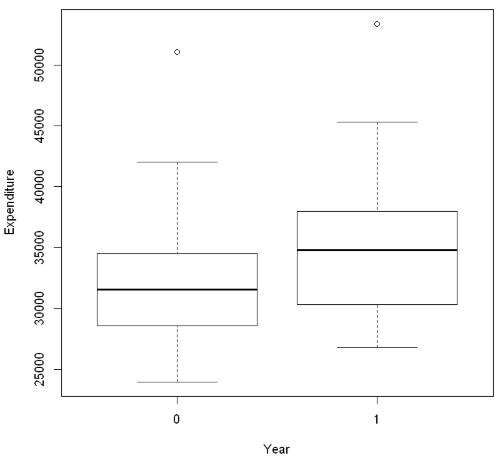
**Hursh Desai**

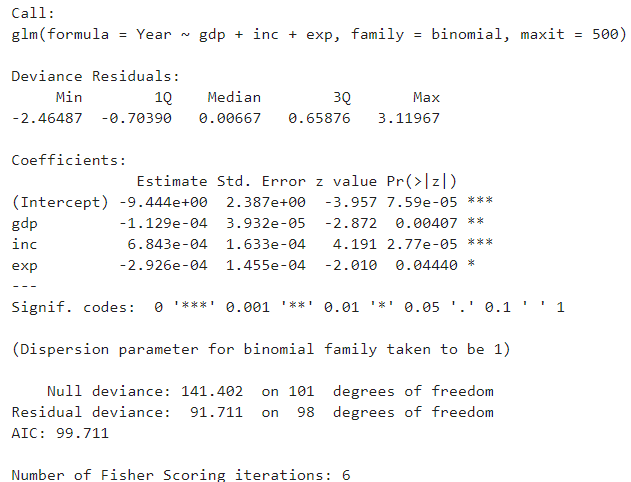
**Question:**

The question I wanted to look at was based off the per capita real gdp, per capita disposable personal income, and per capita personal consumption expenditures of each state in the US could be correctly classify them into the year 2007 and 2012. Coincidentally this question can also be looked upon as very giving somewhat of an answer to the question of if the first term of Obama’s presidency helped the US economy get back on its feet. This question arose from my curiosity of how long it took for America to recover after the 2008 recession. Some cursory searches on the internet revealed many conflicted opinions on this subject some saying that the recession ended in 2009 but the effects can be felt on the economy till this day. I wanted to see the level of association was present between the common economic indicators such as gdp, income, and expenditure and the two years 2007 and 2012. **Analysis:**

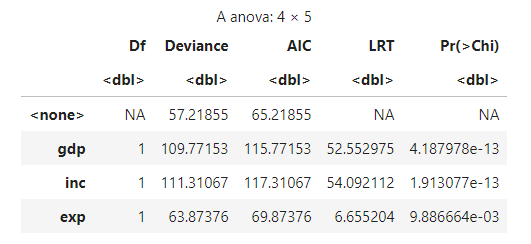
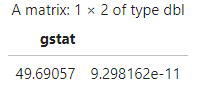
 



Judging based solely off the boxplots there does not seem to be that much separation when it comes to GDP, however for income and expenditure there is a slight increase by 2012 maybe indicating that the economy had gotten back on track by that time. However, D.C. does seem to be an outlier for all three with Connecticut being an additional outlier for Income. We will see if taking them out improves the regression later. This also shows us that the economy has improved even if just slightly since the pre-recession era. However, now the question becomes is it clearly separable from the numbers of 2007.



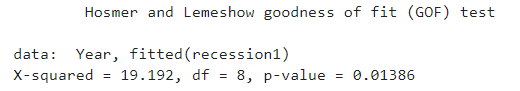
It seems like the regression is statistically significant overall.



The likelihood ratio test also shows that the overall regression is significant. And the likelihood ratio tests for each slope are also highly significant.



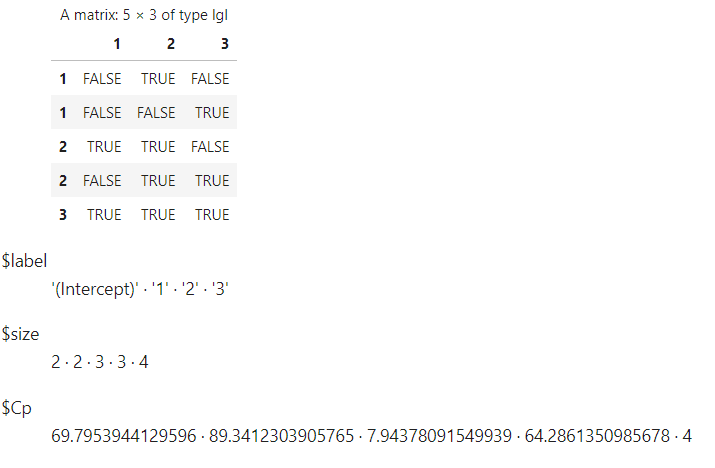
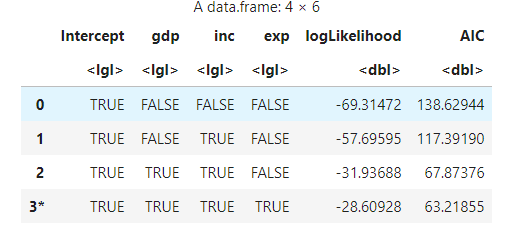
The VIFs are not too high at all meaning that we are okay on collinearity.



The Hosmer-Lemeshow GOF test shows that the data is well fit by this model.



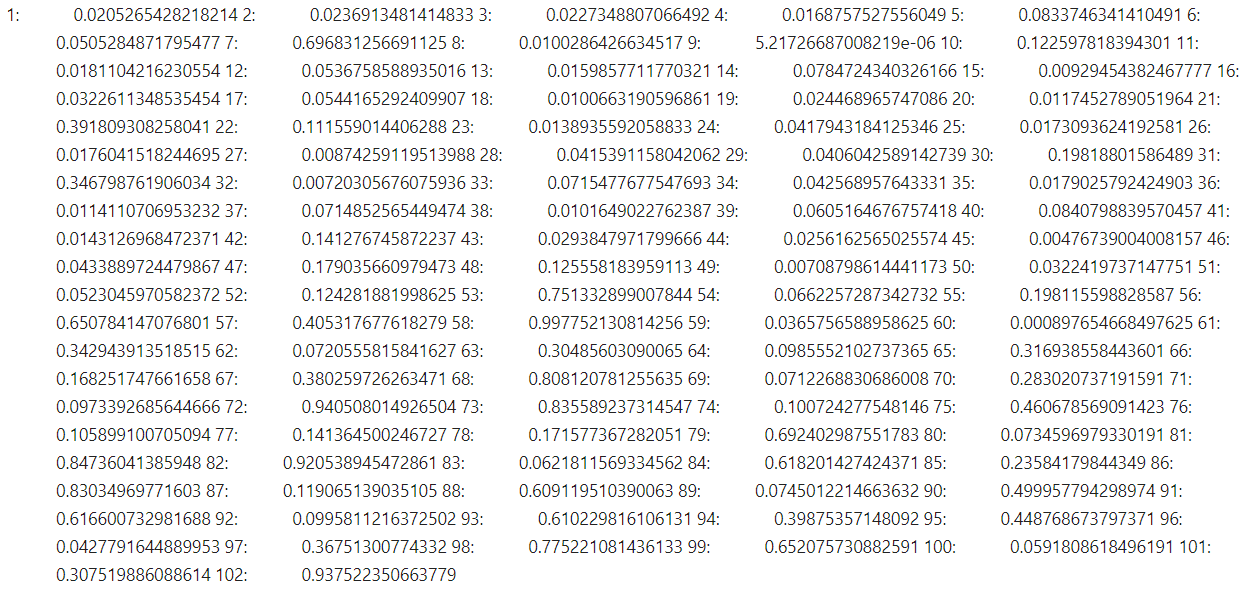
The Somers’ D for this model was a .9006 which means that the proportion of concordant pairs was around 95% and discordant pairs was around 5% which is showing excellent separation.

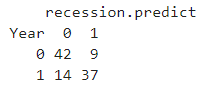
 

Both ways of finding the best subset show that all three predictors are useful in seeing if the two years are seperable.



Going back to the outliers now even though the cook’s distances and residuals are the highest for 60: Washington D.C. there doesn’t seem to be any reason to take them out since our regression here is highly statistically significant. It also doesn’t make sense to take them out since it is necessary to take into account all states to see how the economy fared after the recession. However, if we are to check what happens to the regression if we take Washington D.C. and Connecticut out we see a very little increase in the p-values of the likelihood ratio tests for the predictors and the overall regression and the Somers’ D does not even change in the slightest. The goodness of fit test even shows that the model isn’t fitting the data as well as it did before. 



Looking at the numbers that can be calculated using the Classification Table for the estimate of expeected proportion of observations correctly classified if the actual and predicted memberships were indpendent, we see that if we compare it to the observed correctly classified proportion that we have a good number of correctly classified observations. Thus we can say that using the predictors gdp, expenditure and income, there is significant seperablilty between 2007 and 2012 and thus the economy has recovered since the recession in that given time.