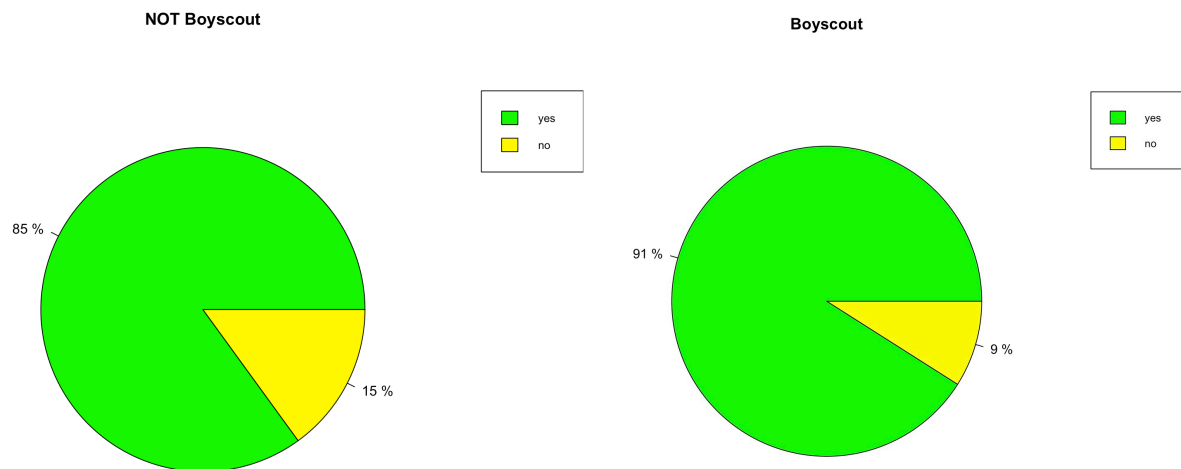# Challenge Problem

Fan Yang (fy2232)

1. Summary

In this project, our goal is to find if there exists any relationship between boy scout status socioeconomic status and their delinquency status. The three variables all have 800 observations. Socioeconomic has 3 levels which are low medium and high. While Boyscout and delinquency both have two levels: yes and no. Since the response variable delinquency status is categorical, I decide to use logistic regression to analyze the relationship.
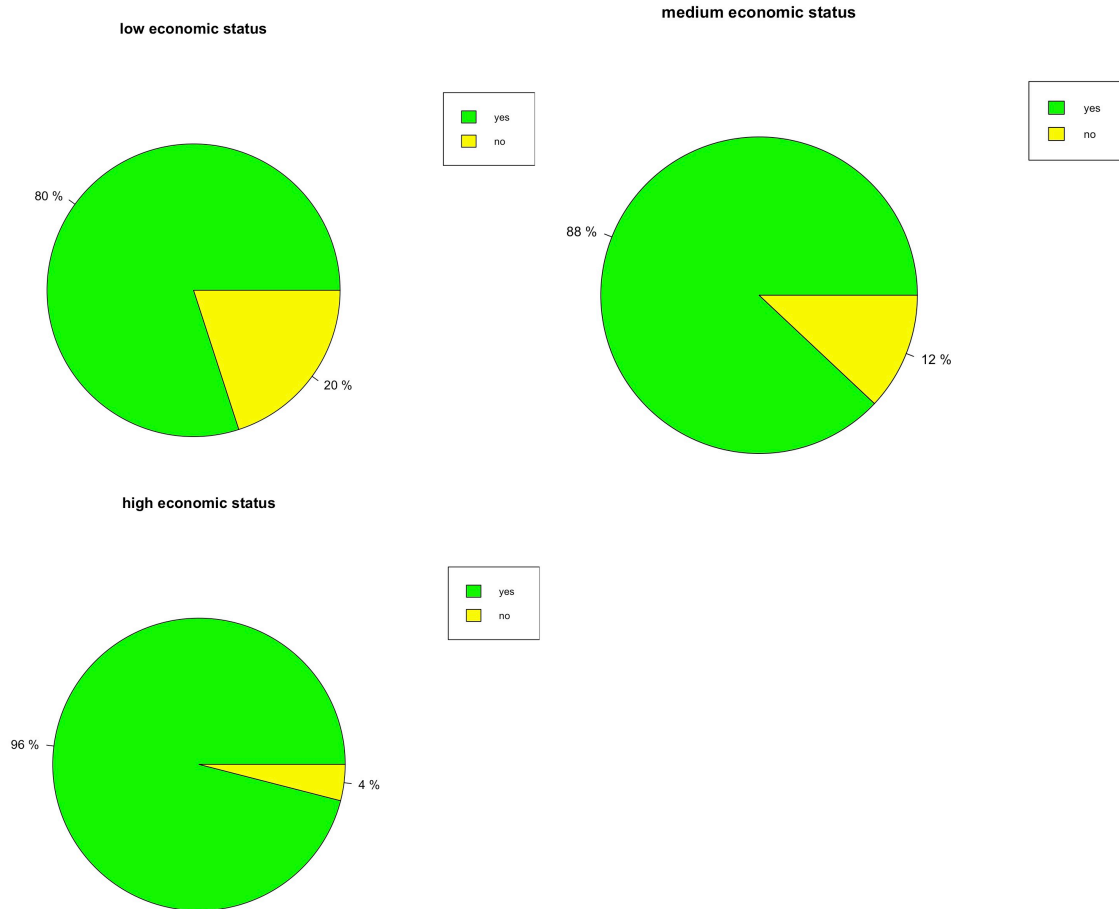
2. Descriptive Statistics

The data set that we use in the analysis consists of 800 individuals cross-classified according to their boy scout status, their socio-economic status and their delinquency status. The boy scout status has three categories (yes and no), the socio-economic status (low, middle, high) and the delinquency status has two categories (yes, no). Fifty percent of these individuals were in boy scouts. Of those who were in the boy scouts 91% were delinquent while 9% were delinquent of those where who were not.

Plot 1. Boyscout status and proportion of delinquency

In addition 31.25% were in the low class, 37.5% were in the middle class and the remaining 31.25% were in the high class. Of those who were in the low class 80% delinquent, while 88% of those in the middle class and insert 96% of the high class were delinquent.

low economic status

medium economic status

yes	
no	

yes	
no	

80 %

20 %

88 %

12 %

high economic status

yes	
no	

96 %

4 %

Plot 2. Economic status and proportion of delinquency

3. Analysis

We use logistic regression to analyze this data and the analysis. In order to get the best model, I used backward AIC method to select variables. For the full model, AIC is 562.33. After removed Boyscout, the AIC has been reduced to 560.33, which is the lowest one. Therefore we can say Boyscout is not significant for our analysis and we should drop it.

Now I refit the model using only economic status as explanatory variable and it is my final model. The estimated odds an individual from low class is delinquent is 6.0 and estimated odds an individual from medium class is delinquent is 3.27. In addition, the analysis shows that we can be 95% that

1. The odds an individual from low class is delinquent is a number between 3.09 and 12.83.

2. The odds an individual from medium class is delinquent is a number between 1.65 and 7.10.

Interpretation of coefficient : When change from low economic status to high economic status, the odds of getting of 1 change by a multiplicative factor equal to 6.0. When change from medium economic status to high economic status, the odds of getting of 1 change by a multiplicative factor equal to 3.27.

4. Conclusion

In the project we studied the any relationship between boy scout status socioeconomic status and their delinquency status. And my final model is a logistic model with response variable as delinquency status and explanatory variable as socioeconomic status.

5. Appendix

Final model:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

$$\pi(\text{delinquency=yes}) = \frac{\exp\{-3.18 + 1.79\text{economicLow} + 1.19\text{economicMedium}\}}{1 + \exp\{-3.18 + 1.79\text{economicLow} + 1.19\text{economicMedium}\}}$$