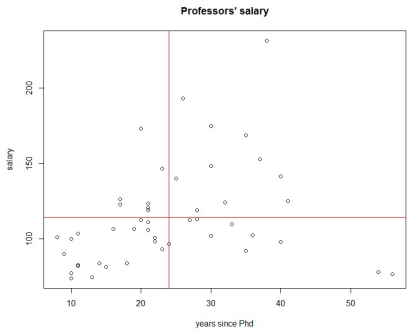
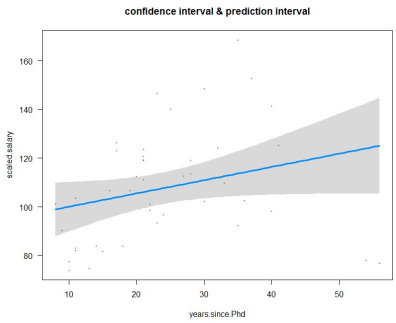
Math 156 Final Project, Mo Pei

Salaries for Professors



We can see a roughly positive relationship. Also, with some outliers and leverage points.

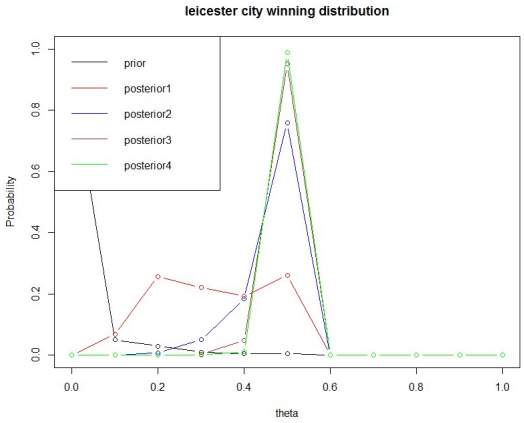
T-confidence interval tells me that the true mean of mean of salary of male full professor is 95% confident between 123592.and 1 130649.5



After remove outliers, the simple linear regression model, inference for years.since.Phd, the coefficient is 0.5443 with p value 0.0615. That means one additional year associated with 0.5443 * 1000 salary increase.

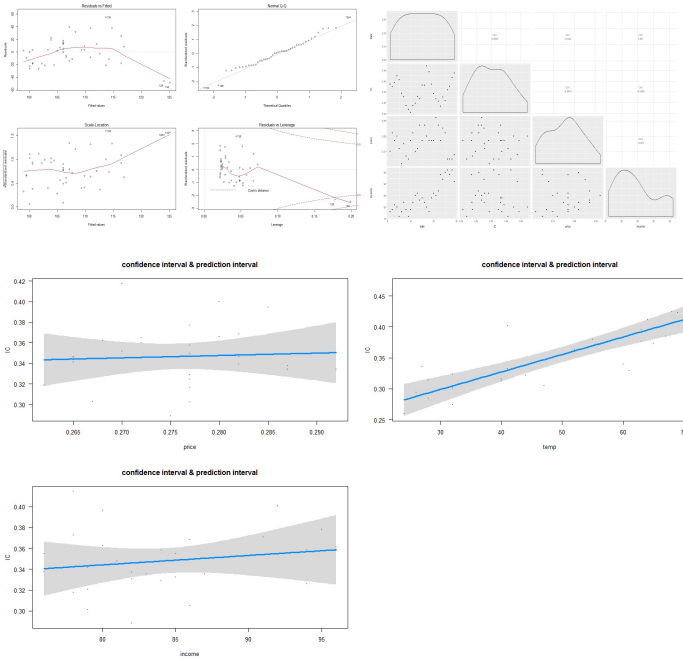
Bayesian statistics

Leicester city soccer team just finished a seasonal year! They, one of worst team last season, just won the premier league championship yesterday! Let see how their winning distribution change. William hill give winning odds of Leicester city 1:5000, which is less than odds of event that Obama will play cricket in England after his presidency. From the plot below, we can see how the winning distribution shift.



Multiple linear regression

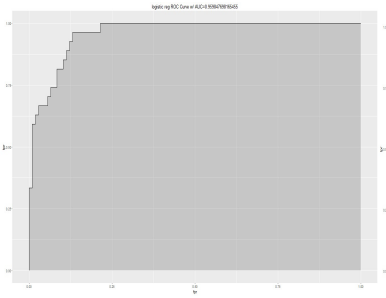
A linear multiple regression model is used to research the relationship between ice-cream consumption and other factors. Among them, I found an interaction term of price and income. After adding, the interaction term, the model becomes better. The interaction term is significant with p value of 0.0491. The overall model is that F test small then 0.01, R-squared: 0.7411, and Adjusted R-squared: 0.698.



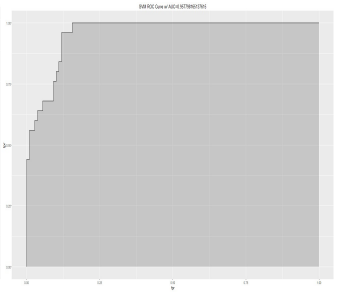
Logistic regression, SVM, KNN, cross validation, machine learning

Logistic regression model and other machine learning models are used to improve classification of labeled customers. I used cross validation, ROC/AUC, confusion matrix, and rmse to evaluate different models and conclude that logistic regression gives the best classifier.

Logistic model ROC/AUC



SVM model ROC/AUC



KNN (n=19)