

Mathematical Statistics Stockholm University Bachelor Thesis **2018:22** http://www.math.su.se

Prediction of a film's profit and its IMDB score

Qi Lin^*

September 2018

Abstract

This thesis aims to find out which factors are associated with the profit of a film as well as its IMDB-score. We use 600 films which have been collected randomly from the website imdb.com. The films was released between 1990 and 2017. We can use Binary logistic regression to study if a film is profitable or not. The first binary logistic regression model is obtained by comparing stepwise procedure with the purposeful selection. The model contains three interactions and five main effects and has a acceptable predictive capacity. The second proportional odds model of IMDB score is obtained by purposeful selection which only includes six main effects. The results show that different predictors variables are correlated with a profitable film and it's IMDB score. These two fitted models have the acceptable predictive capacity with the data set.

^{*}Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: levxy@hotmail.com. Supervisor: Kristoffer Lindensjö and Felix Wahl.