

A Multiple Linear Regression Model Concerning The Swedish Board of Student Finance (CSN)

Yasmin Baghlani*

September 2020

Abstract

This survey study is aimed to detect factors that affect the number of female students that take a loan, at the Swedish Board of Student Finance, by fitting a linear model. Data was abstracted from the National Board of Student Aid (Sweden) and included four independent variables such as education, unemployment, low income (below 60% of the median income) and high income (income above 200% of the median). Furthermore, the dependent variable of the study was the fraction of female loan borrowers from the National Board of Student Aid, in different Swedish municipalities. The main purpose of the study was to examine the effect of each of the independent variables on the fraction of loan recipients and to create a model to predict the percentage of future loan borrowers. We used assumptions of the linear regression model to get a fitted and valid multiple linear regression model, in which we looked at the outliers of the data and checked the assumptions of the regression models. As a result, we fitted a multiple linear regression model from data of 2015 and concluded that the fraction of people with an income less than 60% of the median, in each municipality, was insignificantly correlated with the fraction of loan borrowers. Regarding the significance of the independent variables, the number of educated people in each municipality was the most important variable, which had a positive relationship with the fraction of borrowers. Accordingly, a higher number of educated individuals in each municipality increased the fraction of borrowers. The second most significant independent variable was the fraction of individuals with an income higher than 200% of the median, and it was negatively correlated with the number of borrowers, indicating that the number of wealthy people in each municipality decreased the number of borrowers. The least significant variable affecting the fraction of borrowers was the unemployment rate. The higher the fraction of unemployed individuals in each municipality was, the higher the fraction of borrowers was as well. In order to make use of the regression model, the percentage of loan recipients was predicted using the fitted regression equation on the data from the next year 2016. The results suggest that the predicted values could explain 73.7% of the actual data variations.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: yasminba@kth.se. Supervisor: Ola Hössjer.