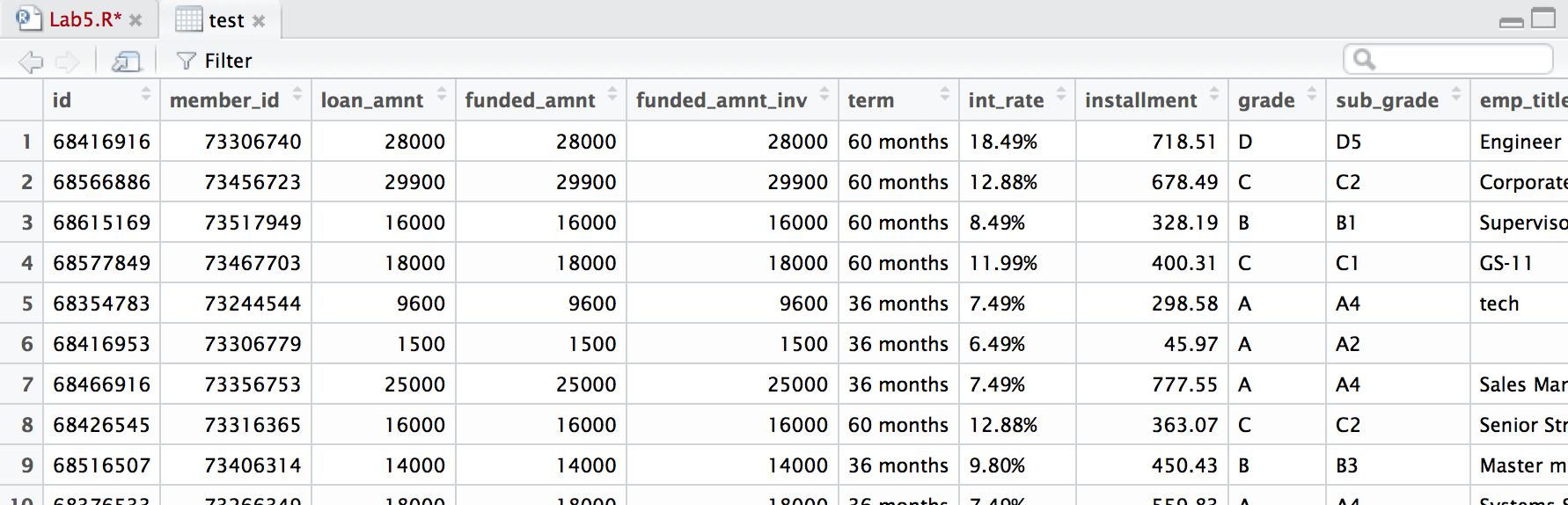


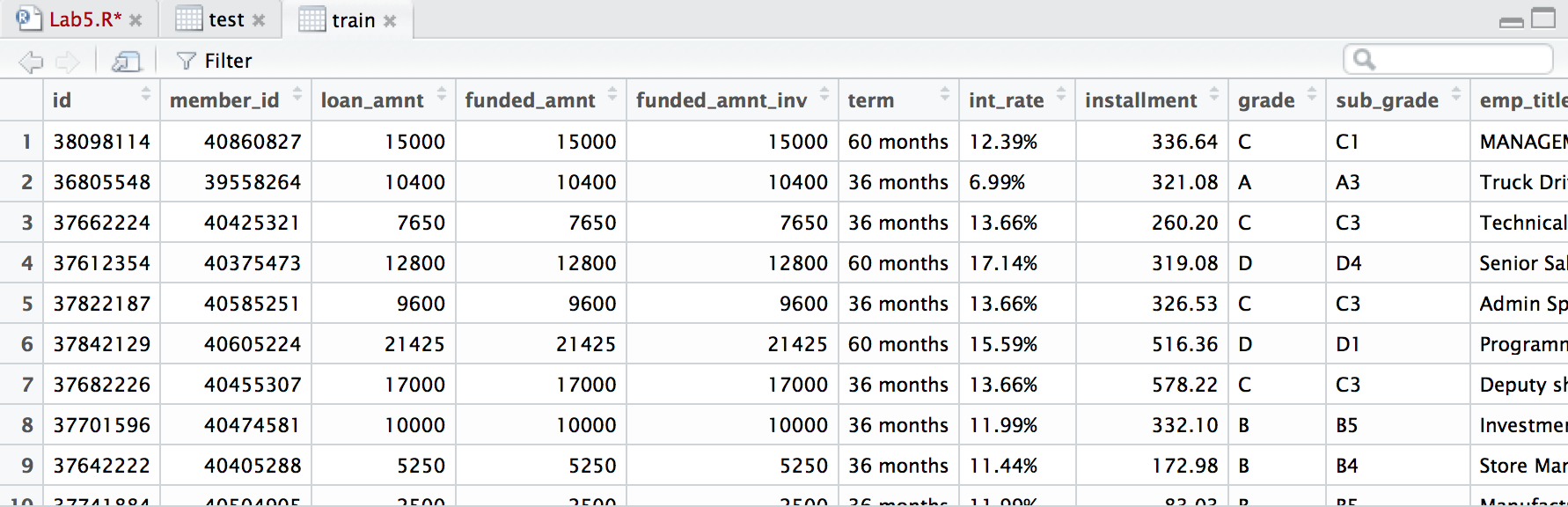
**By: Christian Miljkovic**

**Group: Zachary Fineberg, Suchit Sadan**

**December 16, 2016**

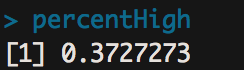
1. Data loading and cleanup.

****

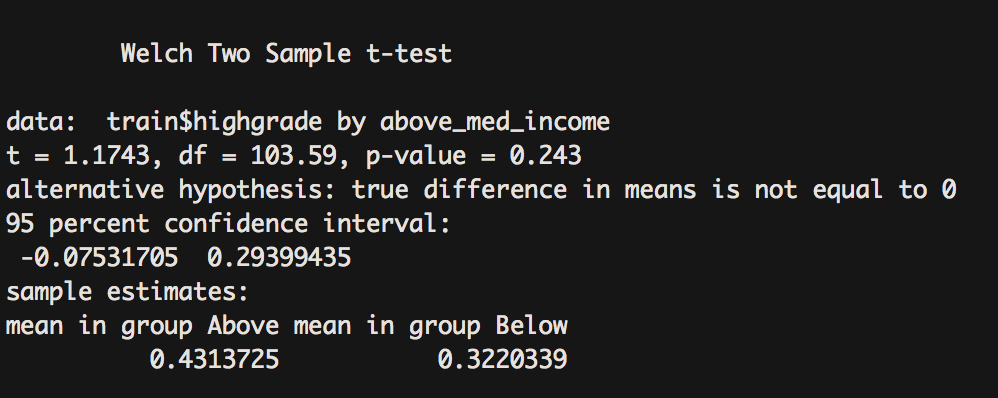
****

## 2. Descriptive statistics.

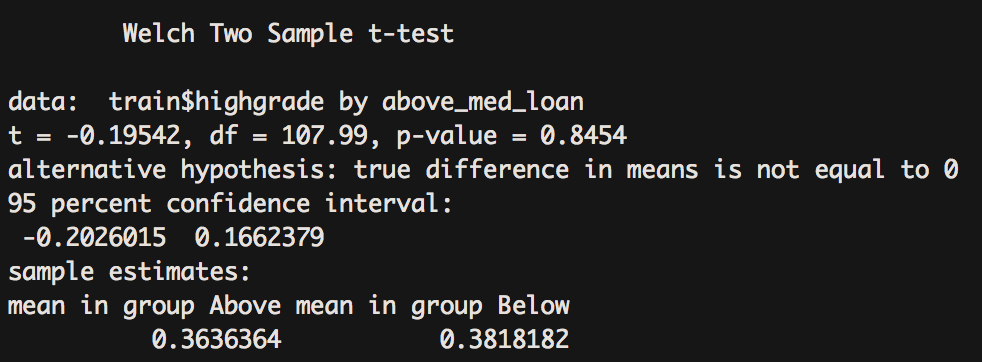
Percent of loans that got high ratings

****

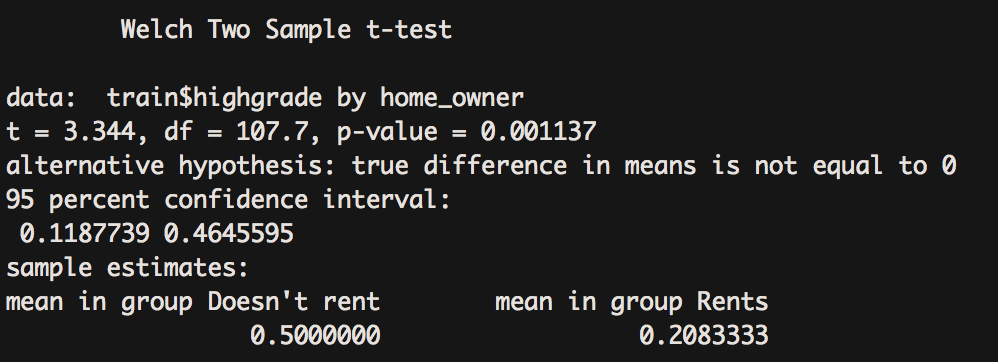
Whether the debtor is above or below the median income level



Whether the loan request is above or below the median loan amount

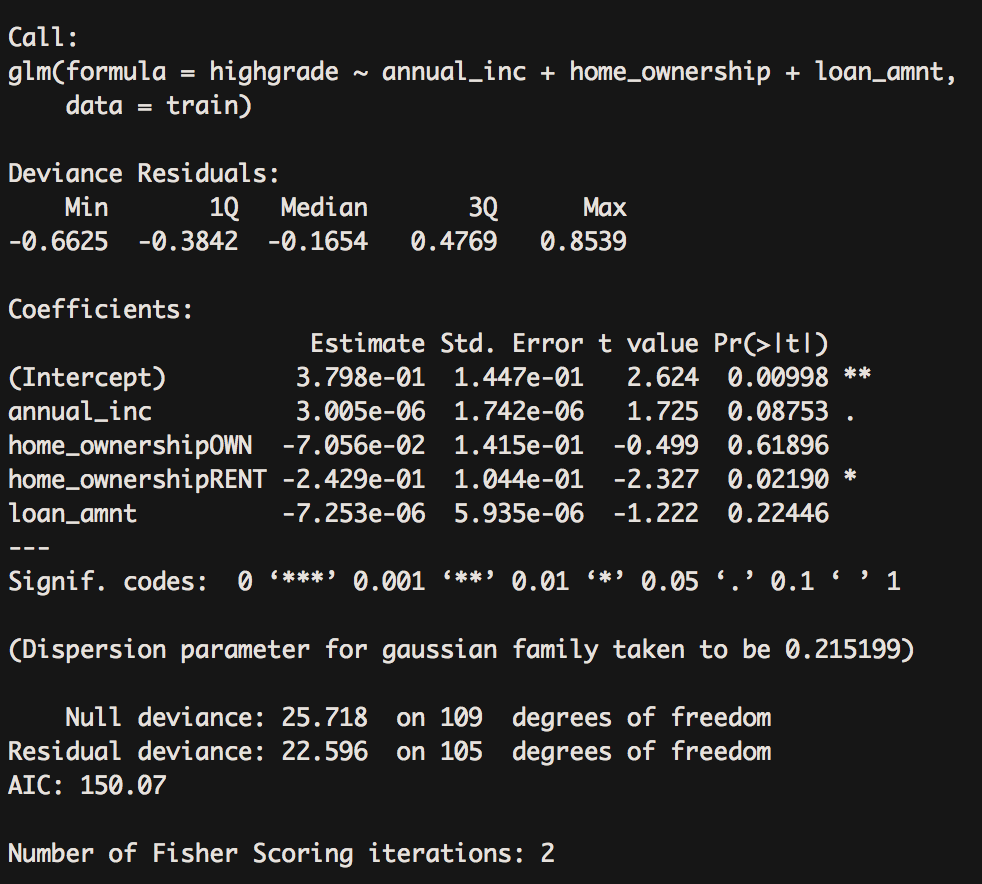


Whether the debtor rents their home or not

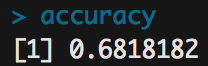
****

## 3. Build a logistic classifier on the training data.

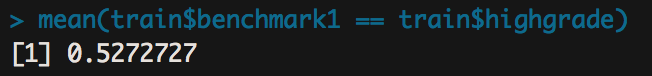
Cut and paste the output produced by the *summary* command.

****

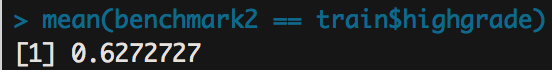
What is the accuracy of this classifier on the training data?



As a benchmark, what would be the accuracy of a classifier that randomly assigns 0 and 1 values as the predicted class?



As another benchmark, what is the accuracy of a classifier that simply assigns a value of 0 to all rows for the predicted class?

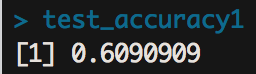
****

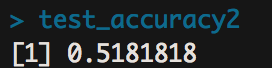
## 4. Supervised learning.

The machine learning based classifier has an accuracy of 0.7545 while the regression based approach has an accuracy of 0.6818

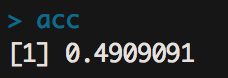
## 5. Model performance on the test data.

Evaluate the accuracy of both of the classifiers you built above (logistic regression + machine learning) on the test data.





As a benchmark, what is the accuracy of a classifier that randomly assigns 0 and 1 values to the test data?



As another benchmark, what is the accuracy of a classifier that simply assigns a value of 0 to all rows of the test data?

