

2. Data

We download the data set **Credit** from online link <http://www-bcf.usc.edu/~gareth/ISL/Credit.csv>, which is provided by the author of the book, “An Introduction to Statistical Learning”. This data set records **Balance**, which is the average credit card debt for a number of individuals, as well as several predictors. The dataset has eleven variables - seven **quantitative** variables, **Income**, **Limit**, **Rating**, **Cards**, **Age**, **Education**, and **Balance**, and four **qualitative** variables, **Gender**, **Student**, **Married**, and **Ethnicity**. Our goal is to understand the relationship between **Balance** and these potential predictors with statistical fitting procedures.

2.1 Pre-modeling Data Processing

In order to fit the regression models, we first preprocess the dataset **Credit** with two steps: * convert factors into dummy variables - which avoids the problem of input data as factors * mean centering and standardization - which provides comparable scales for data analysis
