Data

We first start out with the original data set: credit.csv from the book **An Introduction to Statistical Learning** by James et al. We stored this data set in R as the variable credit. Credit initally contains 12 columns(categories or predictors). However, we remove the first column because it just contains the row numbers of each row, which is irrelevant to our analysis, so we will remember to subset that part out of the data set in R before doing analysis.
There are two important steps in premodeling data proces.

1. convert factors into dummy variables
2. mean centering and standardization

First, we need to factors into dummy variables. We pre-process the data set to make sure that it is in the format that we want. Notable changes are changing categorical variables (such as: M/F, Yes/No, Caucasian/Asian) into 0/1 slots.
Second, we need to standaraize each predictor variables to apply 5 regression linear methods

The first 6 results are shown below :

income	limit	rating	cards	age
-0.8605	-0.4894	-0.4650	-0.6983	-1.2561
1.7253	0.8272	0.8277	0.0310	1.5265
1.6846	1.0135	1.0280	0.7602	0.8889
2.9425	2.0659	2.1074	0.0310	-1.1402
0.3025	0.0699	0.0133	-0.6983	0.7149
0.9920	1.4346	1.3835	0.7602	1.2367

education	gender	student	marriage	asian	caucasian	balance
-0.7839	-1.0343	-0.3329	0.7944	-0.5843	1.0038	-0.4068
0.4960	0.9644	2.9962	0.7944	1.7071	-0.9938	0.8330
-0.7839	-1.0343	-0.3329	-1.2557	1.7071	-0.9938	0.1305
-0.7839	0.9644	-0.3329	-1.2557	1.7071	-0.9938	0.9657
0.8159	-1.0343	-0.3329	0.7944	-0.5843	1.0038	-0.4111
-1.1039	-1.0343	-0.3329	-1.2557	-0.5843	1.0038	1.3724

Table 1: First 6 results of Premodeling data processing results

With this data, we are going to apply five regression methods: OLS(Ordinary Least Squares), RR(ridge regression), LR(lasso regression), PCR(principle components regression) and PLSR(partial least squares regression).