|  |  |  |
| --- | --- | --- |
| **Predictor** | **Expected SOE** | **Rationale** |
| Income | positive | People with higher incomes should usually be able to pay off their credit assuming they didn’t borrow too much. |
| Cards | negative | Having more number of credit cards is not good because it suggests that we tend to borrow more credit. |
| Age | positive | Older people would normally have longer credit histories which could have a positive effect on credit rating. |
| Education | positive | It should not have any effect but People with more number of years education might have better credit ratings because of higher incomes. |
| Gender | Male – positive  Female - negative | The sex of the person by itself should not have any effect on the credit rating but coupled with Income it can have an effect. |
| Student | negative for students | Students might have lower credit ratings because they tend to borrow more as they may be in debt due to tuition loans. |
| Ethnicity | Caucasians - positive African americans - negative Hispanic - negative | Caucasians could have a better credit rating compared to African-Americans and Hispanic, which could be attributed to the systematic racism in the US. |
| Balance | Positive or negative | Balance by itself cannot have any effect but dividing it by limit will show credit utilization which can have an effect. |

1. **What independent (predictor) variables are relevant to predicting credit score? Create a table with three columns for predictor, expected sign of effect, and a one-sentence rationale for effect. Note that all variables provided in the data set may not influence credit score. If some variables are not relevant, explain why they should not influence.**

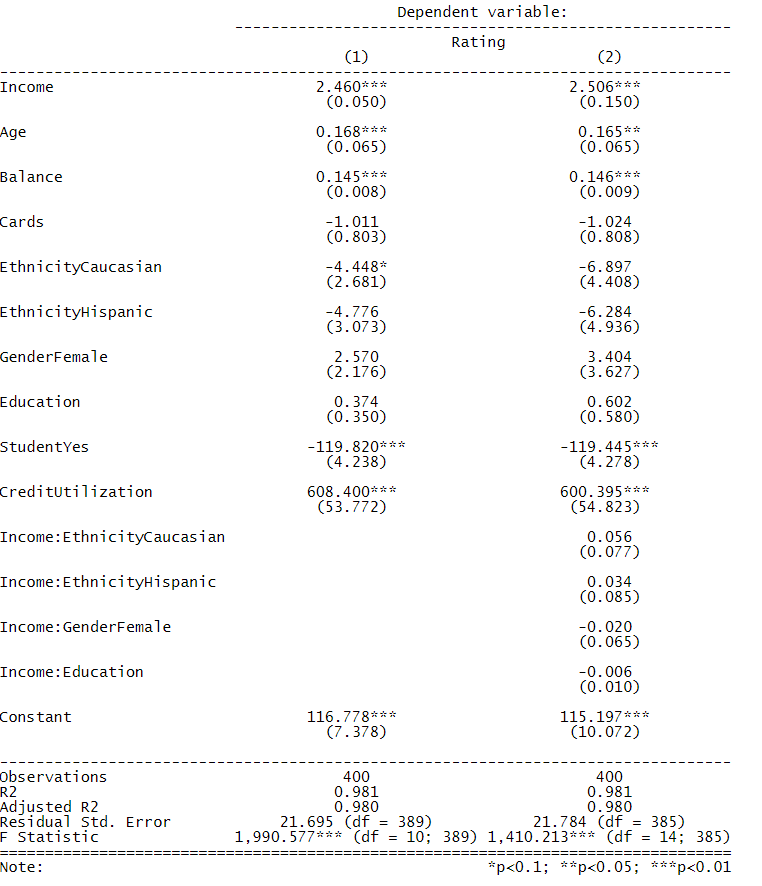
Variables left out

|  |  |  |
| --- | --- | --- |
| Limit | None | We cannot use credit limits as an independent variable as credit ratings determine the credit limit |
| Married | no effect | Being married should not directly have any effect on credit ratings as the ratings do not change because a person is married. |

1. Two models

**Model 1<- lm(Rating ~ Income + Age + Balance + Cards + Education + Ethnicity + Gender + Student + CreditUtilization, data=d)**

**Model 2 <- lm(Rating ~ Income + Age + Balance + Cards + Education + Ethnicity + Gender + Student + CreditUtilization + Income\*Ethnicity + Income\*Gender + Income\*Education, data=d)**



1. Questions
   1. **What is the marginal effect of income on credit score?**

The marginal effect of income is that if income increases by $1000 keeping all other variables constant then the credit ratings would go up by a score of 2.506 in credit ratings.

* 1. **Does having more credit cards help or hurt credit score? By how much?**

Having more credit cards would hurt our credit score. The model also suggests the same ie if we increase the number of credit cards by 1 keeping all other variables constant then our credit ratings goes down by 1.024 in credit ratings.

* 1. **Does maintaining credit card balances help or hurt credit score? By how much?**

Credit balances themselves cannot explain the credit ratings. In the real world, it has been recommended to keep the balance at less than 30% of the credit limit. So, I had added the new feature ‘credit utilization’ which is balance divided by limit. To answer the question, from the model, we can infer that if the average credit card balance goes by $1 then the credit rating goes up by 0.146 in credit ratings. This does not make sense in the real world.

* 1. **Is there a racial bias in credit score, i.e., do African-Americans or Hispanic people have less credit score than Caucasian people, if all other factors are equal? If so, by how much?**

Ethnicity by itself cannot have effect on credit ratings. On doing an interaction plot with race against income, the graph suggested that there is racial bias against African-Americans and Hispanic people in the low income levels. But the model intercepts shows that African-Americans have a credit rating of 115.197 while Caucasians and Hispanics would have 108.3 and 108.1 in credit ratings keeping all others constant.

With the interaction effect, the model suggests that the Caucasians credit ratings over African-Americans improve by 0.056 for higher incomes and Hispanics credit ratings over African-Americans improve by 0.034. This suggests that there is some racial bias at higher levels of income rather than lower levels of income which I had initially hypothesized.

* 1. **Is there a gender bias in credit score, do females have lower or higher credit score than males? If so, by how much.**

From the model, we can infer that females have 3.404 more in credit score ratings over males ie keeping all other factors constant. But after including the interaction term with income, the model suggests that as income increases the credit scores of females goes down by 0.02 in credit ratings. This is true in the real world, as there exists a gender pay gap problem.