1. Create a predictor table (see sample assignment solutions) with three columns for predictor, expected sign of effect, and a one-sentence rationale for effect.

Variable Name (Predictor)	Effect on Credit Rating	Rationale
Income	+	It can indirectly affect a person's credit rating by influencing their creditworthiness, debt-to-income ratio, and ability to make timely payments
Limit	+/-	The higher the credit limit, the higher the utilization which can negatively impact credit rating. It can positively impact if the utilization is low allowing the remaining credit to increase
Cards	+/-	It can indirectly affect a person's credit rating by influencing their credit utilization, payment history, and creditworthiness
Age	+	As age increases, the credit history increases and results in higher credit rating
Education	+	Higher the education increases the income and higher income can increase a person's ability to repay debt and can positively affect their credit rating
Gender		No effect on credit rating
Student	-	Student have low credit history, low limit and limited payment history which have negative impact on credit rating
Married		Marriage status have no impact on credit score
Ethnicity		Ethnicity does not have impact on credit score
Balance	+	As balance increases, credit utilization decreases and credit rating increase

Here target variable is (Y=Rating) which is the measure of credit score rating which commonly ranges from 300 to 850. It is calculated based on various factors such as payment history, credit utilization, length of credit history, and other factors that can indicate how likely a person is to repay their debts on time. Logically, we expect payment history, credit utilization, length of credit history, types of credit cards, types of credit, recent credit enquiries, public records, etc. affect credit rating. Using the literature search, we have considered limit, balance, cards as our initial factors. But income, age, education can indirectly affect credit rating. As income can affect the payment history, credit utilization and credit history directly and these factors are the primary factors affecting credit rating which gives an indirect relation between income and credit rating. To find the Credit utilization, we calculated a column names "creditused" as it has direct affect on credit rating. Similarly, the higher the age, higher the credit history which gives an indirect relation between age and credit rating. Education can also be considered as a factor; the credit score should increase with higher education but the coefficient of the estimated value of education is negative, which is illogical. Hence we dropped the education from the final model.
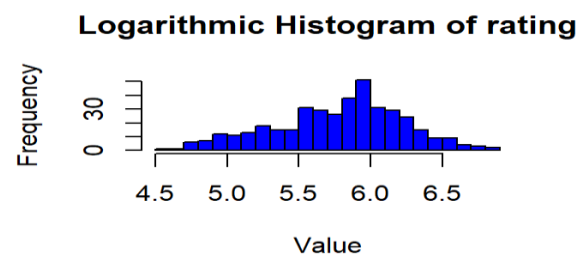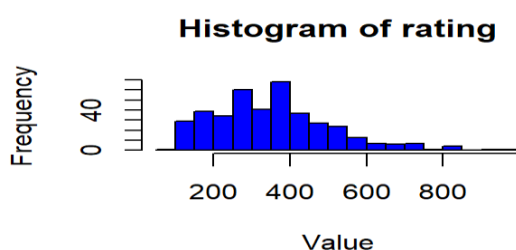
These values show that there in very less skewness in the log curve And the curve with respect to variables is linear.

2. Interpret your models to answer the below question.
   i) What variables predict credit scores and by how much?

| | rating | | | rating | | | rating | | |
Predictors	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
(Intercept)	24.21	18.02 – 30.41	<0.001	31.12	24.28 – 37.97	<0.001	359.28	320.47 – 398.09	<0.001
income	0.09	0.02 – 0.16	0.007	0.09	0.03 – 0.16	0.005			
limit	0.06	0.06 – 0.07	<0.001	0.06	0.06 – 0.07	<0.001			
balance	0.01	0.00 – 0.01	0.013	0.01	0.00 – 0.01	0.010			
cards	7.19	4.60 – 9.77	<0.001	4.68	3.93 – 5.43	<0.001			
age	0.00	-0.06 – 0.06	0.894	0.01	-0.05 – 0.07	0.791			
cards^2	-0.35	-0.71 – -0.00	0.050						
education				-0.23	-0.55 – 0.09	0.159			
gender [Male]				-0.16	-2.16 – 1.85	0.879	-2.92	-33.55 – 27.71	0.852
ethnicity [Caucasian]							-11.56	-49.25 – 26.13	0.547
ethnicity [Hispanic]							-21.95	-65.40 – 21.50	0.321
married [Yes]							13.55	-18.13 – 45.23	0.401
student [Yes]							1.16	-50.09 – 52.41	0.965
Observations	400			400			400		
R^2 / R^2 adjusted	0.996 / 0.996			0.996 / 0.996			0.004 / -0.009		

From the above model, we can interpret that income, limit, balance, cards have positive impact with significant P value. This indicates that all the above-mentioned affecting the credit score. Cards have higher coefficient value and hence we added its second degree which helps in dipping the line at its threshold which is its point at its first derivative = 0

In the other model the R squared = 0.9957 and the RSE = 10.19. but in the above-mentioned model the R squared = 0.9958 and RSE = 10.14. and we have 393 degrees of freedom.

Let us see how the individual models fitting

limit vs rating

income vs rating
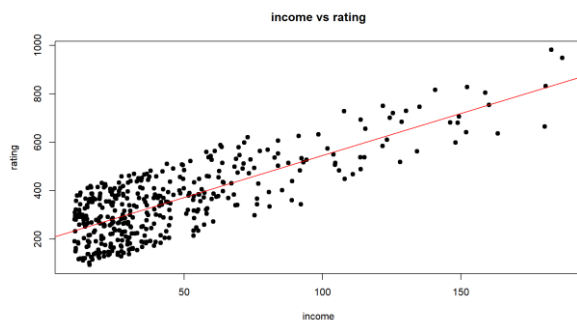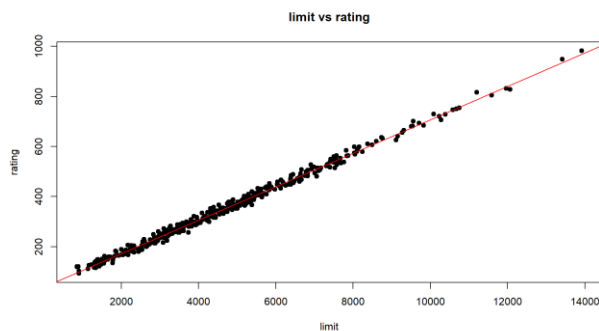
Let us see the linearity of the model

```
lm_secdegree_1=lm(rating~income+limit+balance+cards+age+I(cards^2),data = credit_data)
```


Actuals v. Fitted, price

Normality Plot, rating

Residuals, rating
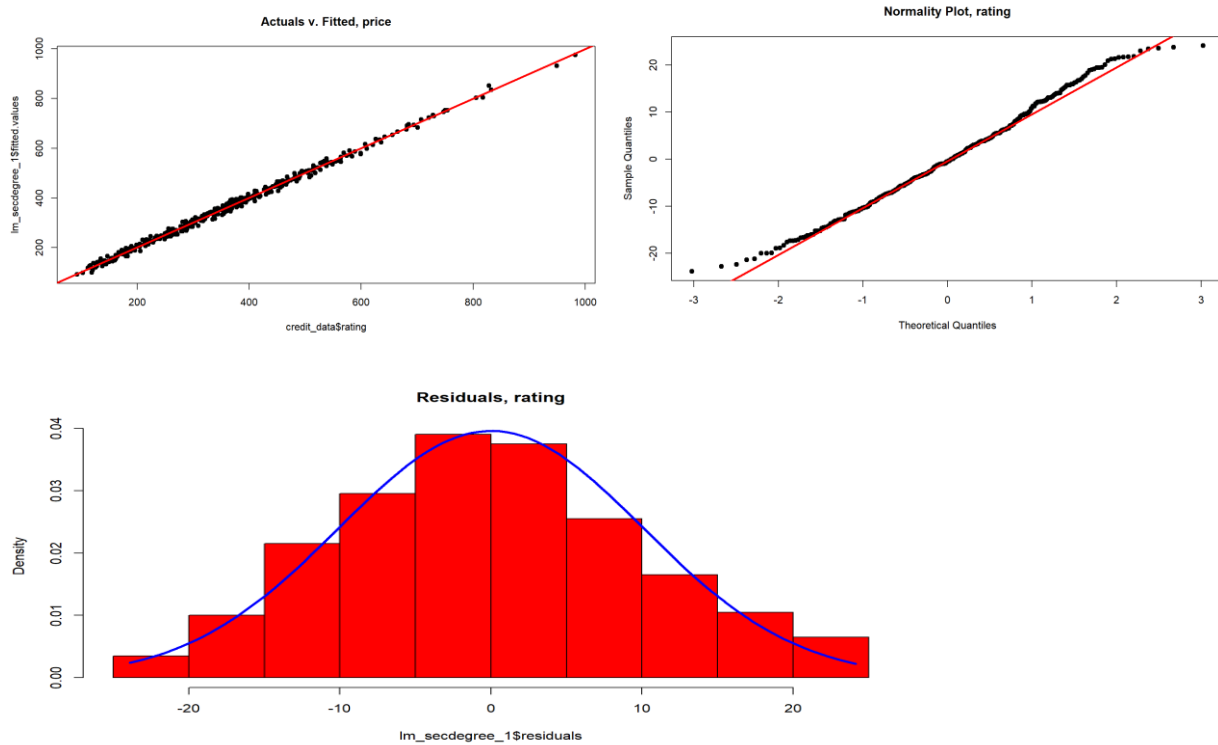
Most of the points are on the line and it is proven to follow linearity.

ii) Is there a racial or gender bias on credit score? If so, by how much?

The mean value of Rating for the Gender = Male is 353.5181
The mean value of Rating for the Gender = Female is 356.2657
The over all mean of rating = 354.94

The difference between them is negligible and hence there is no gender bias on credit score