Homework 1

Please see the instructions about HW submissions on the syllabus posted on canvas. If you are working in a group, you only need to submit one copy per group.

Consider the dataset posted on canvas titled "Credit.csv" where there are many variables. We will only use the following for this assignment:

Rating: Credit card rating of a customer

Income: Annual income of a customer ($)

Age: Age of a customer

Education: Number of years of education of a customer

1-) Explore if a linear relationship is viable between Rating (dependent variable) and all the other variables by obtaining the scatter plots and the respective correlations.

2-) Estimate 3 simple linear regression models between Rating (as the dependent variable) and all the other variables. Report your regression models.

3-) Comment on the intercept and slope coefficients in the context of the problem. What do they represent? Also obtain the 95% confidence intervals for the intercept and slope coefficients for all three models.

4-) Obtain the predictions for a customer rating whose income is 100,000, whose age is 30, whose education level is 13 (for each model separately). In doing so, also obtain 95% and 99% prediction intervals and comment on your findings.

5-) Obtain plots of the actual "Rating" variable versus your three fitted simple linear regression models and the corresponding 95% prediction intervals.

6-) Obtain the R-squared estimates for all three models and using the R-squared estimates obtain the correlations and comment on your findings.

7-) Which model (among the three simple linear regressions) provides the best fit to data? What are the measures you are using to compare the three models?