

Linear Regression Project

This is INDIVIDUAL assignment.

Guidelines:

- a. Write your answer in R Markdown and answer below each question
- b. Attach R Script along with your answer
- c. Show RELEVANT R output in your answers (R Markdown will do this)
- d. Follow the hint given after the question in [] brackets
- e. Read the relevant topics as mentioned in the hint from Ken Black. You can refer any good article/blog from net also (but referring Ken Black is a must)

Task

Refer file **grades.csv**

The school principal wants to build a predictive model for predicting final for his consumption. As a principal he is very keen to have good scores by his students. He has given this data file to you with a request to suggest an appropriate model.

You are required to build at **least 4 models** with different sets of predictors (independent variables). Selection of sets of predictor/s is upon you. Different sets of predictors can be a single variable or more than one variable. However, selection of predictor/s should be based on some logic. For example, for predicting final score of students, roll number cannot be a logical predictor.

You will analyze all 4 models based on following points and recommend **the best** model to the Principal.

1. Describe data of response variable and predictors in terms of key summary statistics like mean, mode, median, standard deviation, range, skewness and kurtosis. Show histogram and box plots also for each variables. Each variable to be explained in 30 words maximum.
2. How predictor/s is related to response variable (final)? [plot scatter diagram followed by correlation test] . Present diagram/s and correlations in the following space. Before diagrams explain relationship in 3 or 4 lines.
3. What are R Square and Adjusted R Square of your final model? Show R Output and explain in 3 or 4 lines.Explain the difference between R Square and Adjusted R Square. Which one is superior and why? Explain in maximum 4 lines.
4. How do you interpret significance value of F^2 statistics? Show R Output. [Fitness of model]
5. Use equation writer of word [Insert → Equation → Insert New Equation and write Regression equation of the best model. Show R Output.
6. What is Durbin Watson Statistics of your model? How DWS is interpreted? Show how do you find dL and dU and design four boundaries in the sample diagram Explore about Durbin Watson Statistics and table from internet. Table is used for finding dL and dU based on which you will design limits. You need to impose your DWS value in the diagram and decide about presence of autocorrelation. Show R Output also.
7. What is VIF for each predictor/s? How do you interpret VIF or what VIF signifies? Max 5 lines. [VIF (Variance Inflation Factor).Show R Output.
8. How do you interpret the significance of slope of predictors based on sig. Value or p-value associated with t -statistics of each predictor/s. Show R output.
9. Test the assumption of Normality and interpret your findings. Show histogram and interpret in maximum 3 lines.
10. Test the assumption of Independent of observations and interpret in maximum 3 lines.

11. Test the assumption of linear relationship and interpret in maximum 3 lines for each predictor. If more than one predictor is used in model then more scatter plots would be required]
12. Test the assumption of Constant Error Variance and interpret in maximum 3 lines
13. What is Standard Error of Estimate of your model and how do you interpret the same. Show with some hypothetical values of predictors. Maximum 300 words
14. Congratulation! You have done a marvellous job indeed and build your first predictive model. I just reminding that regression model is somewhere 50% of a data analyst routine job and has great importance in practical world. Now write a summary of your findings in 250 words which you will show to your reporting manager (before forwarding the model to your client/Principal in this case). This time, no R Output and minimum pictures are needed. Mind it, your reporting manager is a senior statistician/data scientist and do not have time to go into your entire work. He will prefer to read meaningful, to the point and technically correct summary! Here is your chance to impress your boss!
15. This is final stroke! Besides your boss, your client is equally or rather more important to you! Your challenge is this that the Principal/client is not statistics savvy! You need to summarize your work/findings in a non-statistical manner or in a lay man manner and this is indeed challenging. However, no way out and you have to do it in a simple but impressive manner (impressive to client!). Write down summary in 500 words.
16. Now time to show case your work to rest of the world! Prepare a website for the project you created in R. Apply your creativity and make it really impressive. This you must attach with your resume in the shape of giving a link in CV. A worth doing exercise. You may educate your school going wards about individual website and encourage them to show case their projects this way. His/her teacher will be amazed and you will be called by the concerned teacher and head of institution for a thanks giving session!