Regression Analysis

Year 3, Institute of Technology of Cambodia
Department of Applied Mathematics and Statistics

Checklist for Regression Analysis

		Yes	No
Project Scoping	Project description: project description in a short paragraph		
	Project objectives: key objectives for the project		
	Key questions: the macro-level questions the project aim to solve		
	Score - In scope: Outline the work elements that will ultimately allow the team to answer the key project questions - Out of scope: describe the related activities the team will NOT pursue as part of the project		
	Deliveries: what the project output will include		
	Key assumptions and Data request - Factors that are critical for the team's success - Information/data that is needed and how to get it		
	Task flow and members: divide the task for each member		
Data Preparation	Identify response variable and type of model		
	Choose reasonable amount of regressors - This can be based on the objectives/research questions - This can be based on the dataset i.e. correlation between the variable and response variable. - This can be based on the literature reviews. - This can be done by using statistical methods such as **backward selection**, **forward selection** or **stepwise selection**.		
	Identify missing value and decide on remedies		
EDA	Created summary statistics table(s)		
	Confirm if the distribution of continuous variables approximately normal		
	Check the relationship between regressors and response variable		
	Check also the relationship among regressors		

	Identify if there are concerning missing values and outliers		
	Decide on if transformation on variables are necessary - **If necessary, apply transformation such as Box-Cox transformation**		
Statistical Analysis	Identify your model - Check if there are interaction effects between regressors - Identify if there are any confounders (confounding variables) and included them in the model		
	Fit your model - You can fit many models but you may need to present only the final model - **If necessary, apply transformation such as Box-Cox transformation** - **Backward selection**, **forward selection** or **stepwise selection** can also be done here		
	Observe your parameter estimates, p-values and prepare for interpretation		
	Check if there is multicollinearity in your model - This can be done by looking at VIF - If there is, resolve the problem		
	Check if there is influential outliers - This can be done by looking at leverage - If there is, resolve the problem. Another model can be fitted without the influential outliers and then interpret and compare the two models. Leave it to scientists to choose which model they prefer the most.		
	Check all model assumptions and resolve any problems related to your model assumptions		
Discussions and conclusions	Discuss your outputs/parameter estimates and more and compare them with previous research papers if any.		
	State your conclusion on the objectives/research questions		
Limitation and further research	State if there are any concerns in your study that are out of your control		
	Suggest some potential methodologies/remedies that can help to get better model/analysis of the data		

Notation: This checklist is not always perfect. The users may need to think of the nature of the problems they are working on and apply the best one.