Indian Institute of Technology Guwahati
Statistical Inference (MA682)
Project II

Full Marks: 10 Last Date of Submission: **April 28, 2024**

Instructions

- Attempt **all** questions.
- Notations are standard and same as used during the lectures.
- **Write a report in details**. The report should contain necessary calculations, tables, graphs, seed, interpretations and conclusions. There should be **one report file**. Convert it into PDF file and rename as **Project2Group⟨Your group number⟩.pdf**.
- **Prepare codes** (either .R file or .py files). Name the code file as **Project2Group⟨Your group number⟩.R** or **Project2Group⟨Your group number⟩.py**. The code files should be well commented for easy readability. The code file should be standalone file so that all necessary results (numerical, graphical, etc.) can be obtained by running the code file only once.
- **Zip** the report file and code file together. Name the zip file as **Project2Group⟨Your group number⟩.zip**. **One of the group members must send the zip file** to `aganguly@iitg.ac.in` within the deadline mentioned above. This email must be copied to all other group members. The subject of the email must be **[MA682] Project II Submission (Group ⟨Your Group Number⟩)**.
- For example, if your group number is 123, the names of the report file, code file, and zip file would be Project2Group123.pdf, Project2Group123.R (or Project2Group123.py), and Project2Group123.zip, respectively. The subject of the email should be [MA682] Project II Submission (Group 123).
- There will be **no credits** if either appropriate works are not shown in the report and/or error free codes are not submitted.
- You can use inbuilt function or routine for generating random number from any distribution.

Problems

1. (10 points) Find a data-set (from some online source) with one response and at least 10 regressors and register it. The detail steps for registering a data-set is given below. Perform regression analysis by performing point and interval estimation of all parameters, test of significance of the regression, residual analysis, subset selection, finding adjusted R^2, residual analysis, detection of multicolleaniarity, etc. Incorporate appropriate description (source, sample size, description of the variables, etc.) of the data-set in the report. All the steps of the analysis should be mentioned in the report. The results should be interpreted appropriately.

   Steps for Registering Data-set: Each group must use an unique data-set. This is the reason of registering the data-set. After deciding the data-set, one of the group members should send

an email by clicking "reply to all" option of the email carrying this question paper. The body of the email must state the name and source link of the data-set. Once a group choose a data-set and register it, the same data-set must not be used by any other groups. The data must be registered by April 21, 2024 (EOD).

Please follow the instructions very carefully. Last time some of the groups do not follow it properly.