Data Assignment 1

ECO221 (Winter 2023)

Instructions:

1) For all the links in the document, please copy the hyperlink and paste it in your browser. The direct *control + click* approach may not work.
2) Please note that this is a **group assignment**. Please refer to the **attached** Group List.
3) All submissions should contain:
   - A document with the answers.
   - A text (.txt) file containing the R code.
   - The data file generated as a comma-separated-values (.csv) file named as "group_<number>.csv"
4) Please submit the files separately and **not in a zipped folder.**
5) **ONE member** from each group should submit **all the files**.

Overview of the problem

As discussed in the class, the environmental Kuznets curve is a hypothesized relationship between environmental degradation and per capita income such that, in the early stages of economic growth, pollution emissions increase and environmental quality declines due to increased production. However, beyond a threshold level of per capita income, higher economic growth is associated with environmental improvements due to increased demand for better environmental quality and/or a shift in production towards 'cleaner' products. Naturally, for any demand to materialize into policy action - requires political intervention which is often a function of *who* is demanding *what* outcomes. Hence, the power and/or income distribution, i.e., inequality also likely contributes to the relationship between income and environmental quality (Torras and Boyce 1998). For more details, read the slides shared on Google Classroom and the main reference shared below:

Torras, M. and Boyce, J. (1998) Income, Inequality, and Pollution: A Reassessment of the Environmental Kuznets Curve. Ecological Economics, 25, 147-160. PDF linked **here**.

1. As a first step, pick *one* environmental quality measure from the sheet linked **here**.
2. Upon downloading the original data for all years from the National Data Analytics Platform (NDAP), you will notice that environmental quality *measure* of your choice has multiple *indicators*. Note further that, the yearly data may not be at a district-level and in fact may be available at a finer granularity. Transform this original dataset into a district-year level dataset that includes a unique district-year ID for each row in the sample.
3. Next, please merge the district-year level environmental quality data with the corresponding state-year wise economic output data, i.e., the net state domestic product (SDP) at constant prices (shared **here**) provided by the Reserve Bank of India accessed on the Database for the Indian Economy (DBIE) portal.

4. Finally, merge your dataset with the district-level Gini index from the following paper by Mohanty et al. (2016). PDF linked **here**.
5. Prepare detailed summary statistics for all the variables (tables; histogram; box-plot; shape of the distribution, skew). Are there any outliers?
6. Now, using this data, estimate the following regression for *any one* environmental quality indicator of your choice. Summarize the results in a table **and interpret in plain English**. Note that i indexes districts, t indexes years, and $u_{i,t}$ is random error.

$$\boldsymbol{Environmental\ Quality\ Indicator}\ (\boldsymbol{EQI})_{i,t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 SDP_{i,t}} + \boldsymbol{u_{i,t}}$$

7. Visualize the model residuals (i.e.,$\hat{u}_{i,t}$) on a plot having the environmental quality indicator on Y-axis and SDP on the X-axis. Now, construct a second plot having $\hat{u}_{i,t}$ on Y-axis and SDP on x-axis. Finally, construct a third plot having predicted values of the environmental quality indicator on Y-axis and true values of the environmental quality indicator on X-axis. How are these three plots related, if at all? Explain.

8. Plot a histogram of $\hat{u}_{i,t}$ and verify that $\sum_{i,t} \hat{u}_{i,t} = 0$.
9. Finally, estimate the following regression, summarize the results in a table and interpret. Note that i indexes districts, t indexes years, and $\gamma_{i,t}$ is random error.

$$\boldsymbol{EQI_{i,t} = \alpha_0 + \alpha_1 SDP_{i,t} + \alpha_2 SDP^2_{i.t} + \alpha_3 SDP^3_{i.t} + \alpha_4 GINI_i + \gamma_{i,t}}$$