Data Assignment 2:

Recall from class that we are interested in understanding the impact of agricultural growth on health outcomes. **Please check the dependent variable assigned to your group. This is the dependent variable you will be working with for your class project**. Note that the analysis must be carried out separately for each of the two seasons: Kharif and Rabi.

1. **Model Specification:**
   a. Model the assigned health indicator as a function of the set of independent variables for which you have the data (i.e., GDP, Number of Beds, etc.). Also, you are encouraged to include additional explanatory variable(s) that might explain the variation in the health indicator. For some examples see below:
      - https://www.spring-nutrition.org/sites/default/files/understanding_the_linkages_between_agriculture_and_health-ifpri_2006.pdf

      In addition, please explore alternative model specification(s) and justify your choice(s). Pick the most appropriate model based on your judgement and carry out the following analysis in parts (b – e).
   b. Visualize the model residuals (i.e., $\widehat{u_{i,t}}$) on a plot having the health indicator on y-axis and yield index on x-axis. Now, construct a second plot having $\widehat{u_{i,t}}$ on y-axis and yield index on x-axis. Now, construct a third plot having predicted values of the health indicator on y-axis and true values of the health indictor on x-axis. How are these three plots related, if at all? Explain.
   c. Plot a histogram of $\widehat{u_{i,t}}$ and verify that $\sum_{i,t} \widehat{u_{i,t}} = 0$.
   d. Plot a histogram of $\widehat{u_{i,t}} * x_{i,t}$ and verify that $\sum_{i,t} \widehat{u_{i,t}} x_{i,t} = 0$

2. **Monte Carlo Simulations:**
   a. Regress the assigned health indicator on the yield index and obtain the intercept and slope coefficient estimates.
   b. Treat these estimates obtained as your true population parameters and carry out the Monte-Carlo simulation procedure to comment on the consistency of OLS estimates.

   Hint: Construct a sub-sample by randomly discarding 20% of your original dataset and work with the remaining random sample (i.e., 80% of your data) to run the MC simulations.

3. **Dummy Variables:**
   a. Pick the most appropriate model in question 1 and introduce dummy variables for state-groups.
   Example:
   $$H_{i,t} = \beta_0 + \beta_k \overrightarrow{X_{k_{i,t}}} + \beta_{south}.D_{i,south} + u_{i,t} \quad (1)$$
   Here, i indexes districts and t represents time. $H_{i,t}$ is the assigned health indicator. $\overrightarrow{X_{k_{i,t}}}$ is the set of k explanatory variables (i.e., GDP, Number of Beds, etc.), $D_{i,south}$ is a dummy variable which takes value 1 if district i lies in lies in the southern group of states (see classification below), and is 0 otherwise. β_k is the slope coefficient associated with the k^{th} variable and β_{south} is the coefficient

associated with the dummy variable for southern states. Clearly, $\hat{\beta}_{south}$ represents the difference in average health outcome level between the southern states and the rest of the country. Therefore, $\hat{\beta}_{south}$ accounts for a structural break in the mean outcome level across designated state-groups.

b. Use the t-statistic to test the null that there is no such structural break across different state-groups.
c. Can you interpret equation as an F-test to verify whether there is a structural break across southern and non-southern state-groups? (Hint: Refer to class notes on full and restricted model based hypothesis testing).

(For class project: creative component example): Do you expect the impact of agricultural growth on health outcomes to change over time? If yes, how would the regression model account for such a difference? Justify.

State Classification by Geographical Zones:

North Zone: Himachal Pradesh, Punjab, Uttarakhand, Uttar Pradesh and Haryana.

East Zone: Bihar, Orissa, Jharkhand, and West Bengal.

West Zone: Rajasthan, Gujarat, Goa and Maharashtra.

South Zone: Andhra Pradesh, Telangana, Karnataka, Kerala and Tamil Nadu

Central Zone: Madhya Pradesh and Chhattisgarh

North East Zone: Assam, Sikkim, Nagaland, Meghalaya, Manipur, Mizoram, Tripura and Arunachal Pradesh

Alternate State Classification by Agroecological Zones:

https://vikaspedia.in/agriculture/crop-production/weather-information/agro-climatic-zones-in-india