**Indian Institute of Technology Kharagpur**

**Department of Industrial and Systems Engineering**

**Statistical Learning Lab (IM39202)**

**Topic:** Non-linear modeling, Polynomial regression, Splines and GAM

**Objective:** In this lab the students will learn the basic techniques of fitting the non-linear regression model. Also, how to choose the best degree polynomial fit. It also includes splines and GAM regression model fit.

**Tools required:** R and R-studio.

**Tasks:**

1) Import the designated data file and Display first few rows of the dataset.

2) Data cleaning and pre-processing and Perform preliminary analysis to show how the variables are related to each other. Use scatter plot, box plot etc. to visualize how different variables impact the response variable.

3) Convert categorical inputs or consider it while fitting the data

4) Fit a linear model first

5) Then fit polynomial regression model

6) Analyze the fitted model using ANOVA

7) Select best fit degree polynomial

8) Fit spline with varying knots and GAM model.

**Report Format:**

Create an R markdown file and submit the pdf with all the code snippets, plots, results and explanation.

**Dataset: (Go to link download the zip file and export one of the below data for above operations)**

1. Poverty data (<https://online.stat.psu.edu/stat501/lesson/1>)
2. Iqsize data (<https://online.stat.psu.edu/stat501/lesson/5>)