

**Department of Statistics**  
**Savitribai Phule Pune University**  
**ST-205 - Practical 1: Simple Linear Regression**

- Q.1. In the data set contained in the file titled 'GroundWater.txt', X gives pH of the well water and Y gives Bicarbonate (parts per million) of the well water. The data is given for a random sample of wells chosen from a small town near Pune. Answer the following questions:
- (a) Is there a linear relationship between X and Y ? Justify your answer. Justify the statistical tool/method applied by you.
  - (b) Fit a linear regression model to Y taking X as predictor. What is your conclusion about (a) based on this model?
- Q.2. Refer to the data set in the file titled 'MathScores.txt'. A group of 7 students is given a particular performance enhancing drug. The students appear for a mathematics test after the consumption of the drug. The scores are recorded in the first column and the drug concentrations are provided in the second column.
- (a) Fit a simple linear regression model to the 'MathScore' taking 'DrugConcentration' as the predictor variable.
  - (b) Give 90% and 95% confidence intervals for the intercept,  $\beta_0$  and the slope,  $\beta_1$ .
  - (c) Examine normality of the errors.
  - (d) Also check the independence of the errors.
  - (e) Compute the R<sup>2</sup>. Comment on it.
  - (f) Give an overall conclusion about the model.
- Q.3. Consider the data set in the file titled 'AutoInsurance.txt'. The column X gives the number of claims and Y gives the total payment for all the claims (in thousands).
- (a) Fit a simple linear regression model to Y taking X as the predictor variable.
  - (b) Find a 95% prediction interval for the mean total payment if the number of claims is known to be 50.

- (c) Give a 95% prediction interval for the total payment if the number of claims is known to be 50.
- (d) If the average number of claims is increased by 100, then the insurance company states that the average total payment of claims increases by 300 thousands. Apply appropriate test procedure to examine this statement at 1% level of significance.
- (e) Examine normality of the errors.
- (f) Check the assumption of the constant variance of the errors.
- (g) Also check the independence of the errors.
- (h) Compute the  $R^2$ . Comment on it.
- (i) Give an overall conclusion about the model.