Indian Institute of Technology Bombay

SI 422: Regression Analysis

Problem set 1

January 22, 2020

Three sets of paired samples are given for the following variables:

(1) The purity of oxygen produced by a fractionation process and the percentage of hydrocarbons in the main condensor of the processing unit,

(2) The weight and systolic blood pressure of a group of males in the age group 25-30,

(3) The number of pounds of steam used per month at a plant and the average monthly ambient temperature.

Answer the following questions for each data sets.

(i) Draw the scatter plot.

(ii) Fit a simple linear regression model and draw the fitted regression line in the scatter plot.

(iii) Report the fitted values and corresponding residuals.

(iv) Calculate the coefficient of determination and comment.

(v) Test whether intercept of the above model is zero.

(vi) Test whether slope of the above model is zero.

(vii) Construct analysis-of-variance table and comment.

(viii) Find 95% confidence interval on the intercept.

(ix) Find 95% confidence interval on the slope.

(x) Find 95% confidence interval on the mean purity, symbolic blood pressure and steam usage when the hydrocarbon percentage, weight of a male in the age group 25-30 and average monthly ambient temperature are 1, 160 and 58 unit respectively.

(xi) Find 95% prediction interval on the purity, symbolic blood pressure and steam usage when the hydrocarbon percentage, weight of a male in the age group 25-30 and average monthly ambient temperature are 1, 160 and 58 unit respectively.

(xii) Compute and plot 95% confidence and prediction bands around the fitted line.

(xiii) Which is the wider band and why?