Homework 4 -Tuo Wang:

1). This program analyze data in one.txt file using the least squares regression method and the least absolute deviation method. Using each, the program assigns a line of best fit to the data in the X column to best predict the data in the y column. The program takes the residuals of the methods estimators for mean and median to determine which wins. In this case, the OLS has a lower residual for the mean estimator – almost 0 – and the LAD has lower residual for the median estimator.

2) As a regression equation cannot be formed with all the data in the population set since population are theoretically too big, the data which is sampled to make up the regression line is resampled multiple times, but always in samples of the same size. And the relevant values of each sample are determined. Thus a virtual histogram of the distribution of each value is made, and with this average, a more accurate estimate can be made of how much the selected varies in the population.