**Beyond ‘Expectation’ : Analyzing Cab fares through a Quantile Regression approach**   **Name: Anubhav Hazra**  **Roll number:466 Supervisor’s Name:Prof. Rahul Roy**

**SYNOPSIS**

**Introduction**

This is a study of fares of cab rides, and various factors that influence the fare in different ways,i.e. when the fares are usually low, the relative importance of some predictor might be more, and vice versa.

**Objective**

Ordinary least squares regression gives us the average effect, or more specifically the conditional mean of the response,given the predictor values.We would want to get an idea of the corresponding conditional quantiles.This would sketch out a better picture of the varying distributions, alongside relaxing the classical model’s homoscedasticity assumption.We would have separate regression lines for each quantile,whose parameter estimates would be obtained by minimizing the sum of skewed(weighted) absolute deviations from the response values,where the weights are a function of the particular quantile we want to fit.

**Data Description**

The dataset, from Kaggle.com, consists of data on taxi rides for New York City of 2018,with 3120 rows and 7 columns, namely– distance, duration, extra, tolltax, day, time and the fare(response).

**Methodology**

Let q];the qth quantile of a random variable y is that value of y below which q proportion of the observations lie.The quantile regression problem can be tackled by minimizing the following loss function w.r.t :

**= +**

Where **,** i.e.the model error, k is any positive real number, is the ith row of the design matrix X, is the solution vector for the qth regression quantile,and the conditional quantile estimate is = X .

**Expected Result**

We would monitor the change in these parameter estimates, corresponding to the different covariates, over different deciles(i.e.10% quantiles),allowing a comparative analysis of the effectiveness among the predictors and also with the estimates given by least squares regression.