Zagat’s publishes restaurant ratings for various locations in the United States. The file RESTRATE.xls contains the Zagat rating for food, décor, service, and the price per person for a sample of 53 restaurants located in New York City and 53 restaurants located in Long Island. You want to develop a regression model to predict the price per person based on a variable that represents the sum of the ratings for food, décor, and service and to write a report on your findings. The

1. State your statistic objective for the data set.

The main objective of this is lab is to analyze the effect of the multiple variables food, décor, service, and location have on the price per person.

1. Perform exploratory data analysis (EDM), such as numerical measures or the box-and-whisker plot for variables on the data set.
2. Construct a scatter diagram of price against summated rating. Describe a relationship that you may see. Does this appear to have some association (linear or non-linear)?
3. Assuming a linear equation, use the least-squares method to compute the regression coefficients b0 and b1. State the regression equation for predicting the cost per person based on the summated rating.
4. Interpret the meaning of the Y intercept b0, and the slope b1.
5. Construct a scatter plot of price against each of the variables food, décor, and service separately. Describe a relationship that you see from each diagram. Does any of these appear to have association?
6. From (a) and (b), does any simple linear model appear to hold? Run some testing including a LINE analysis to substantiate why or why not.
7. Does any multiple regression model appear to hold? Run some testing to substantiate why or why not. If so, find the regression equation to predict price from location.
8. Develop a regression model to predict the price per person based on a variable that represents the sum of the ratings for food, décor, and service, and on location (New York City (Locate = 0) or Long Island (Locate = 1)).

Is this regression significant? Report the results of the appropriate test, and interpret its meaning (in plain English).

Are any other variables (Neighborhood or Cuisine) use for the regression analysis? For example, when you classify Cuisine as Asian, American, or Others, can you use them as dummy variables?

1. Does summated rating have significant impact on price, following adjustment for location? In particular, are New York City’s restaurants significantly more expensive or significantly less on average than those in Long Island?
2. Consider an interaction term in the model. Then at the 0.05 level of significant, determine whether it makes a significant contribution to the model.
3. Summarize and comment on your results and an applicability of your model equation(s).
4. Predict the cost of per person for a restaurant with a summated rating of 50.
5. Compute the coefficient of determination r2, and interpret its meaning.
6. Perform the residual analysis on your results and determine the adequacy of the fit of the model.
7. At the 0.05 level of significant, does evidence exist of a linear relationship between the price per person and the summated rating?
8. Construct a 95% confidence interval estimate of the mean cost per person for all restaurants with a summated rating of 50.
9. How useful do you think the summated rating is as a predictor of cost? Explain.