R·I·T	**Rochester Institute of Technology** **Golisano College of Computing and Information Sciences** **Department of Information Sciences and Technology**

Lab 2 (4 points)
Linear Regression

This lab consists of two parts, which use linear regression for prediction and statistical interference.

Part I

This question involves the use of multiple linear regression on the Auto data set.

(a) Produce a scatterplot matrix, which includes all of the variables in the data set.
(b) Compute the matrix of correlations between the variables using the function cor() . You will need to exclude the name variable, which is qualitative.
(c) Use the lm() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summary() function to print the results. Comment on the output. For instance:
    i. Is there a relationship between the predictors and the response?
    ii. Which predictors appear to have a statistically significant relationship to the response?
    iii. What does the coefficient for the year variable suggest?
(d) Use the plot() function to produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?
(e) Use the * and : symbols to fit linear regression models with interaction effects. Do any interactions appear to be statistically significant?
(f) Try a few different transformations of the variables, such as log(X), $X^{0.5}$, X^2. Comment on your findings.

Part II

This question should be answered using the Carseats data set, which is part of the ISLR package.

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.
(b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!
(c) Write out the model in equation form, being careful to handle the qualitative variables properly.
(d) For which of the predictors can you reject the null hypothesis $\mathbf{H}0: \beta j = 0$?
(e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.
(f) How well do the models in (a) and (e) fit the data?
(g) Using the model from (e), obtain 95% confidence intervals for the coefficient(s).