

Multiple Linear Regression

Q1: MLR with 2 or more variables – Machine tools company

Benedrix, a machine tool company is interested in understanding the impact of machine hours and production run, on its overhead cost.

The data on a monthly basis for 3 years is provided in the dataset:

MLR_FactoryOverhead.csv

Fit the regression equations:

- Overhead = F (machine hours)
- Overhead = F (production runs)
- Overhead = F (machine hours, production runs)

- 1) Find the R-Squared in all 3 cases.
- 2) How would you explain the additional lift in R-squared of the combined model vs each of the individual model?

Q2. MLR with categorical variables – Courier Service

GoKart is a regional delivery service company providing different types of package delivery services. An analyst wants to estimate the cost of shipping a package as function of cargo type.

Cost of 15 randomly chosen packages of approximately the same weight, shipped to almost similar distance, is provided in the file: ***MLR_CourierService.csv***

- 1) Estimate the appropriate multiple linear regression equation to predict the cost of shipping a package.
- 2) Provide interpretation for the regression coefficients.
- 3) Which cargo type is the costliest? Which one is least costly?
- 4) How well does the regression fit the sample data? How can the goodness of fit be improved?
- 5) Predict the cost of shipping a package with semi-fragile cargo.

Q3: MLR with multiple variables – Employee Salary

An HR analyst in Unitech Pvt Ltd, wants to predict the annual salaries of given employees using the potential explanatory variables in the file **MLR_EmpSalary.csv**

- 1) Estimate the appropriate multiple linear regression equation to predict the salary of an Unitech employee using all explanatory variables.
- 2) Do we need to exclude certain columns? Why?
- 3) Which department employees are paid the highest? By how much?
- 4) Do you see any discrimination in salaries earned by male and female employees?
- 5) What would be the estimated salary of a Data Scientist (joining engineering) with 10 years of work experience. This woman has 18 years of total education, and will be supervising 4 junior employees.

Q4: MLR with Interaction effect – Stock Price Prediction

Stock market analyst are continually looking for reliable predictors of stock price. Consider the problem of modelling the stock price of Utility companies. Two variable that are thought to influence the stock price(Y) are return of average equity (ROE) and annual dividend rate.

Data for 16 utility stocks are provided in the file **MLR_StockPrice.csv**

- 1) Estimate the MLR equation from the given dataset
- 2) Interpret the Adjusted R-Squared value and each of the coefficients
- 3) Revise the Stock Price prediction model for utility companies, to include and interaction term between ROE and Annual dividend rate. Does this new model fit the data better than was done in Problem 10.

All datasets are available here: <https://github.com/Accelerate-AI/Data-Science-Global-Bootcamp/tree/main/ClassAssignment/Assignment06>