

## Assignment-2 (Multiple Linear Regression)

1. Suppose we want to determine how two variables influence sales at stores in a chain of women's apparel stores (annually, in dollars per square foot of retail space). A regression that describes these sales would be useful for choosing new locations for stores as well as evaluating current franchises. Each of the 65 stores in the data is located in a shopping mall and occupies 3,000 square feet. One explanatory variable is the median household income of the area served by the mall (in thousands of dollars); the other is the number of competing apparel stores in the same mall. We expect outlets in malls that serve more affluent neighborhoods will have higher sales, whereas we expect outlets that face more competition will have lower sales

Answer the following questions:

- Draw the scatter plot matrix and find the correlation matrix. Comment on the association between the pairs of variables?
- Run a linear regression model of sales per sq. ft on income and number of competitors. Find the summary statistics (OLS estimates of the regression coefficients, p-values, confidence intervals, R-squared, Adjusted R-squared, Root mean squared error) and interpret it.
- Compare the marginal and partial slopes, and interpret it by drawing a path diagram.
- Suppose that several high-profile technology businesses move their corporate headquarters to locations near a mall, boosting incomes in that area by \$5,000. If the collection of retailers in the mall remains the same, how much would you expect sales to increase at an outlet of the type considered in the data? Should you use the marginal or partial slope for income?
- Carry out an F-test for testing the hypotheses  $\beta_1 = 0$  and  $\beta_2 = 0$  simultaneously. This is an omnibus test for no regression.
- Suppose F-test leads to the conclusion of no regression but a particular regression coefficient shows significant result. Should we consider it significant?