

Answer these two questions and provide a report in PDF format.

Q1

A national insurance organization wanted to study the consumption pattern of cigarettes in all 50 states and the District Columbia. The variables chosen for the study are given in ex1 text file. The states are given in alphabetical order.

- (a) Test the hypothesis that the variable Female is not needed in the regression equation relating Sales to the six predictor variables.
- (b) Test the hypothesis that the variables Female and HS are not needed in the above regression equation.
- (c) Compute the 95% confidence interval for the true regression coefficient of the variable Income.
- (d) What percentage of the variation in Sales can be accounted for when Income is removed from the above regression equation? Explain.
- (e) What percentage of the variation in Sales can be accounted for by the three variables: Price, Age, and Income? Explain.
- (f) What percentage of the variation in Sales that can be accounted for by the variable Income, when Sales is regressed on only Income? Explain.

Q2

The Education Expenditure data given in ex2 text file. The measured variables for the 50 states are:

Y - Per capita expenditure on public education

X 1 - Per capita personal income

X 2 - Number of residents per thousand under 18 years of age

X3 - Number of people per thousand residing in urban areas

The variable Region is a categorical variable representing geographical regions:

1 = Northeast, 2 = North Central, 3 = South, 4 = West

Analyze the effects of regional characteristics on the regression relationship.