

ERSH 8320, Fall 2007

Homework #2

Due: Thursday 9/27/2007 at 11:59:59pm

Homework format: all homework answers must be provided electronically. Please answer each question with as much detail as you can provide. Do not provide direct SPSS (or other package) output for any question (the exception being figures/graphs).

The data (on the proceeding pages) were collected by Christopher Bingham, in a study for the *American Almanac for 1974*, with the exception of the fuel consumption variable, which came from the 1974 *World Almanac*.

The variables collected are:

TAX – 1972 motor fuel tax rate, in cents per gallon.

INC – 1972 per capita income in thousands of dollars.

ROAD – 1971 thousands of miles of federal-aid primary highways.

DLIC – Percent of population with driver's licenses.

FUEL – Motor fuel consumption in gallons per person.

For this homework assignment, we will be trying to predict FUEL based on the remaining variables. Using a Type I error rate of (0.05) for all hypothesis tests, please answer the following questions (without providing SPSS output)

1. Run a multiple regression analysis using the TAX, INC, and DLIC variables to predict FUEL.
 - a. For each parameter of the regression, report the parameter estimate and standard error of the estimate. Report the percent of variation accounted for by the regression. Provide an interpretation of these values.
 - b. Conduct an omnibus hypothesis test of all regression parameters Report the hypothesis being tested, the test statistic, and the p- value. What do you conclude about the test? Provide an interpretation of the result.
 - c. For each slope, conduct a hypothesis test that tests whether the parameter is non-zero in value. Report the hypothesis being tested, the test statistic, and the p-value. What do you conclude about the test? Provide an interpretation of the result.
 - d. Provide a scatterplot of the standardized residuals (as the y-axis) with the predicted values. Can you detect any trends in the plot? Are there any potential outliers?
 - e. Report the observation with the largest Cook's D, Leverage, DF Beta, and Standardized DF Beta (along with the values – there will be multiple for the Beta

statistics). Does one observation have the largest of each of these values? If so, try to develop an explanation of why this observation is unlike the rest.

- f. Delete the observation with the largest Leverage and re-run the regression. Report any differences in parameter estimates. Are there large changes to the results? Are there any new outlying data points?