Midterm Exam

Regression Modeling

Your professor (once) drove to work each morning on the dreaded Long Island Distressway, from a faraway suburb (that sits high on a hill, overlooking things green and beautiful...) The amount of time the drive took each day is relatively constant, but it did vary systematically with some obvious factors and with some less obvious factors. These included:

- Time of departure from home - there was a distinct peaking effect.

- Day of the week. Monday seemed to be the peak day, Friday was less crowded.

- Holiday. Holidays were special.

- Rain. Actual occurrence of rain had an obvious effect. 99.9% of New Yorkers cannot drive in the rain. Your professor could. So could all drivers of SUV's. They were always willing to demonstrate this.

- Snow. Snow on the road had a surprising effect on drive time. Many peopled stay home. New Yorkers are also unable to drive in snow. However, unlike rain, they do not like even to try to drive in snow.

- The idiot effect. An idiot in an SUV who weaves their way into (up) a tree or into some other car affects everyone's time for most of the morning.

The data set midterm.dat contains a large sample of observations on these effects and the drive time. Your assignment is to formulate a regression model for drive time which incorporates these observed effects. Some notes about the data set:

- Drive time is coded in minutes

- Exit time is coded in minutes past 6:30 in the morning. E.g., 5.0 means 6:35.

- Day is coded 1 for Mon., 2 for Tues, etc.

- Holiday, Rain, Snow, and Idiot are ordinary dummy variables.

Tasks:

1. Estimate the coefficients of your model.

2. Identify which are the most significant effects on drive time, and which are the least important.

3. Compute the predicted drive times based on your equation.

4. Note that only one day was affected by the idiot effect. Compare the actual and fitted values for this day, and comment.

5. Compare the day of the week effects. Is the drive typically worse on Tuesday than on Thursday?

6. Test the hypothesis that my guess about weekday effects is incorrect. That is, test the hypothesis that variation in drive time is not explained by day of the week.

7. Test the hypothesis that the expected drive time on a rainy holiday is the same as on an ordinary day on which it does not rain.

8. My worst day took a miserable 195 minutes to get to work. What were the conditions that day. What drive time would your regression model predict for that day?