Fall 2000 Course 513-697: Applied Linear Models

Assignment 9. Hand in on Monday November 6

---

1 NWNW4 Problems 7.7, 7.8, 7.10, 7.20, 7.21, 7.23, 7.27

2 Consider the list of variables in the "1993 cars" dataset in www.epi.mcgill.ca/hanley/c678/ .

a On a-priori grounds, list 5 variables (including at least one categorical variable) that you think would most influence the fuel economy of the cars listed in the dataset.

Take "a-priori" to mean "before running any regressions or peeking at the relationships in the data" -- i.e. base your list on physical considerations.

b Last year, when I asked students for variables that would predict the fuel economy, some listed "Fuel tank capacity". Why is this variable not "eligible" this year?

c List other variables that might have been "eligible" under last year's assignment but are not according to this year's wording. Explain the reasons for each such variable.

d From your ?limited/?extensive knowledge of cars --and without looking at relationships in the data -- predict the rank order of the 5 variables in (a) as for their influence on fuel economy.

- *For now, do not consider interaction terms or higher order models.*

- *Be explicit about your measure of "degree of influence", and how you will measure the influence of the categorical, vis-a-vis the interval, variables.*

- *Make your list 2 ways ...*

(i) when each of the 5 is considered alone

(ii) with the 5 ordered according in decreasing influence: imagine the best "single variable" regression model [from (i)], then add to this the best "second variable", etc...

e OPTIONAL
Use the dataset to test the accuracy of your predictions in 2d*(i)*.
[no marks deducted for poor predictions.. this is not a course in engineering]