Department of Statistics

COURSE STATS 330

Assignment 3, 2006

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 14th September.

Question 1

The data for this question relate to educational expenditure in the United States. For each of the 50 states, we have data on the following variables:

expend: Current expenditure per pupil in average daily attendance in public elementary and secondary schools, 1994-95 (in thousands of dollars)

ratio: Average pupil/teacher ratio in public elementary and secondary schools, Fall 1994

salary: Estimated average annual salary of teachers in public elementary and secondary schools, 1994-95 (in thousands of dollars)

takers: Percentage of all eligible students taking the SAT, 1994-95

total: Average total score on the SAT, 1994-95 (response)

Note the SAT (Scholastic Aptitude Test) is a standard test sat by US high school students in their final year of high school, and is used to evaluate applicants seeking admission to US universities.

The question here is to what extent student performance (the SAT score) is associated with the other variables.

(a) Read the data into R, checking the data as usual. The data are in the file **`sat.csv`** on the course web page. [5 marks]

(b) You are required to fit a suitable model to these data, taking note of the following points:

- Are all the variables required in the regression? Use variable selection techniques to choose a suitable subset if not. For your chosen subset:
- Is the relationship between the response and the explanatory variables linear? If not, can a suitable transformation be made?
- Are there any outliers or influential points in the subset?

- Is the normality assumption satisfied?
- Is the equal variance assumption satisfied?

If you see any problems take appropriate corrective action until the model is satisfactory. [15 marks]

(c) Use your final model to carefully explain the relationship between the explanatory variables and the response. In particular, comment on the following:

- Do these data provide any evidence that, other things being equal, higher teacher salaries are associated with better student performance?
- Other things being equal, is student achievement associated with the percentage of students taking the SAT? If so, can you explain why this might be so?

[5 marks]

Question 2.

(a) In Question 1 in Assignment 2, we treated the variables **`exposure, danger`** and **`predation`** as numeric variables. Would the model have been improved if they had been treated as factors? [5 marks]

(b) Use stepwise model selection to select a model, treating **`exposure, danger`** and **`predation`** as factors. Do you wind up with the same model you chose in Assignment 2? [5 marks]

(c) Repeat the prediction you made in assignment 2, using the model fitted in (b). Do you get a more convincing prediction? [5 marks]

Data for Question 1

	expend	ratio	salary	takers	total
Alabama	4.405	17.2	31.144	8	1029
Alaska	8.963	17.6	47.951	47	934
Arizona	4.778	19.3	32.175	27	944
Arkansas	4.459	17.1	28.934	6	1005
California	4.992	24.0	41.078	45	902
Colorado	5.443	18.4	34.571	29	980
Connecticut	8.817	14.4	50.045	81	908
Delaware	7.030	16.6	39.076	68	897
Florida	5.718	19.1	32.588	48	889
Georgia	5.193	16.3	32.291	65	854
Hawaii	6.078	17.9	38.518	57	889
Idaho	4.210	19.1	29.783	15	979
Illinois	6.136	17.3	39.431	13	1048
Indiana	5.826	17.5	36.785	58	882
Iowa	5.483	15.8	31.511	5	1099

Kansas	5.817	15.1	34.652	9	1060
Kentucky	5.217	17.0	32.257	11	999
Louisiana	4.761	16.8	26.461	9	1021
Maine	6.428	13.8	31.972	68	896
Maryland	7.245	17.0	40.661	64	909
Massachusetts	7.287	14.8	40.795	80	907
Michigan	6.994	20.1	41.895	11	1033
Minnesota	6.000	17.5	35.948	9	1085
Mississippi	4.080	17.5	26.818	4	1036
Missouri	5.383	15.5	31.189	9	1045
Montana	5.692	16.3	28.785	21	1009
Nebraska	5.935	14.5	30.922	9	1050
Nevada	5.160	18.7	34.836	30	917
New Hampshire	5.859	15.6	34.720	70	935
New Jersey	9.774	13.8	46.087	70	898
New Mexico	4.586	17.2	28.493	11	1015
New York	9.623	15.2	47.612	74	892
North Carolina	5.077	16.2	30.793	60	865
North Dakota	4.775	15.3	26.327	5	1107
Ohio	6.162	16.6	36.802	23	975
Oklahoma	4.845	15.5	28.172	9	1027
Oregon	6.436	19.9	38.555	51	947
Pennsylvania	7.109	17.1	44.510	70	880
Rhode Island	7.469	14.7	40.729	70	888
South Carolina	4.797	16.4	30.279	58	844
South Dakota	4.775	14.4	25.994	5	1068
Tennessee	4.388	18.6	32.477	12	1040
Texas	5.222	15.7	31.223	47	893
Utah	3.656	24.3	29.082	4	1076
Vermont	6.750	13.8	35.406	68	901
Virginia	5.327	14.6	33.987	65	896
Washington	5.906	20.2	36.151	48	937
West Virginia	6.107	14.8	31.944	17	932
Wisconsin	6.930	15.9	37.746	9	1073
Wyoming	6.160	14.9	31.285	10	1001