

Department of Statistics

COURSE STATS 330

Assignment 3, 2008

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 18th September.

In this assignment, you are required to develop a prediction formula, after possible selection of variables. The data for the assignment relate to salaries of the presidents of certain colleges in the United States. The purpose of developing a formula is to assist in the setting of salaries in the future.

Q1. The file **presidents.txt** on the course web site is a text file containing data on several US liberal arts colleges. The following variables are recorded:

Comp: The compensation (salary plus benefits) of the current president (\$000)

Gen: Gender of current president (0=female, 1=male)

Rev: College revenue (\$m)

Endow: Size of college endowment (\$m)

Top50: Is college ranked in top 50 liberal arts colleges? (1=Yes, 0=No)

Net: College's net assets (\$)

GS: Grants and supports (Fundraising, \$m)

PExp: Program expenses (cost of instruction, \$m)

TExp: Total expenses (Budget, \$m)

In this assignment, you are required to develop a prediction equation to predict the president's compensation, given the values of the other variables. To give the assignment a little added spice, there will be a prize for the best prediction, based on the prediction error for an actual college. In the event of a draw, the winner will be chosen at random from those having the equal best prediction.

Specifically:

- (a) Read the data into R, checking the data as usual. The data are in the file **college.txt** which is on the course web page. [5 marks]
- (b) Fit a model to the data using all the explanatory variables. Is it better to use `comp` as a response or `log(comp)`? [10 marks]
- (c) Using the response you chose in (b), develop a prediction equation that will allow you to predict the salary of a college president. Are all the variables required in the regression? Use “all possible regressions” variable selection techniques to choose a suitable subset if not. If you see any problems with your chosen model take appropriate corrective action until the model is satisfactory. [20 marks]
- (d) Use your final model to predict the compensation of the female president of a top 50 college having revenue of \$117m, an endowment of \$156.256m, net assets of \$235.154m, grants and supports of \$20.676m, program expenses of \$73.357m and total expenses of \$90.712m. Express your prediction as a point prediction. [5 marks]

Q2 (STATS 762 only).

How well would stepwise regression do in the college presidents example? To investigate this, assume that the model chosen in Q1 is correct. Then,

- (a) Generate 1000 data sets by using the complete set of X variables, but generating a new set of responses each time by adding to the fitted values from the Q1 model a vector of random normal errors. Set the standard deviation of the errors to be equal to the estimate obtained in Q1.
- (b) For each data set, perform a stepwise regression and record the model chosen.
- (c) Make a table of the 1000 chosen models. What fraction of these coincide with the true model? What do you conclude about stepwise regression as a technique?

R hints: You will need to construct a loop for the 1000 repetitions. Inside the loop, you will need to use `step` to select the subset. To record the subsets chosen, make a character vector **result** of length 1000 before entering the loop. Within the loop, assign the result of the `step` function to a variable **stuff** say. You can pull the chosen model out of the object **stuff** and assign it to the *i*th element of **result** by the using the (unfortunately complex) code

```
result = numeric(1000)
```

```

for(i in 1:1000){

Lines to generate data, select model, assign to stuff

result[i] = paste(sort(attr(stuff$terms,"term.labels")),
collapse="+")

}

```

Use the function **table** to tabulate how many times each model is chosen.

Data for Question 1 follow:

Comp	Gen	Rev	Endow	Top50	Net	GS	PExp	TExp
173.096	1	26	37.845	0	70.133	3.393	22.874	24.272
261.630	0	64	346.894	0	382.635	13.620	25.259	48.566
173.903	1	26	51.281	0	65.067	3.011	15.000	21.688
309.540	1	26	153.740	0	224.182	13.767	33.203	41.696
122.350	1	51	35.709	0	40.355	7.140	37.519	47.019
282.188	1	94	113.871	0	191.081	33.809	55.580	60.538
340.650	1	232	890.511	1	1247.966	66.621	74.036	100.281
241.399	1	69	82.476	0	122.134	13.662	52.786	59.561
245.408	1	47	90.595	0	131.230	10.522	37.886	43.654
351.613	1	46	114.529	0	191.029	13.257	42.192	46.577
591.006	1	104	433.244	1	599.771	24.803	86.360	95.670
238.763	1	34	42.432	0	68.651	4.371	28.241	31.656
251.846	0	119	433.854	1	625.437	36.997	76.000	90.878
225.150	1	208	428.878	1	629.827	30.056	118.123	133.112
281.004	1	39	543.456	1	645.713	27.190	78.349	96.261
218.325	1	44	138.529	1	178.014	4.370	32.303	38.843
262.182	0	32	57.949	0	85.153	8.614	21.788	27.251
262.797	0	15	367.185	1	457.696	14.118	47.798	56.805
184.947	1	30	40.661	0	84.550	7.825	30.536	34.393
259.576	1	125	353.383	1	500.765	14.456	58.538	71.755
299.929	1	138	464.535	1	604.690	14.936	101.946	115.040
203.021	1	89	201.459	0	308.387	25.893	70.984	75.038
224.107	0	99	390.129	1	517.520	22.202	55.826	65.060
190.715	1	44	61.950	0	110.159	13.746	30.397	32.778
300.588	1	118	318.218	1	487.897	44.804	43.418	62.601
258.708	1	117	430.467	0	534.045	30.624	73.058	81.556
268.586	1	139	449.509	1	572.209	42.033	69.431	90.285
247.800	1	81	158.255	1	202.304	12.268	83.927	88.747
157.155	1	62	206.534	0	251.455	17.863	64.585	75.464
180.400	1	76	353.361	0	245.259	10.927	33.464	40.943
239.900	1	75	266.509	1	329.815	11.486	56.917	74.558
253.532	1	100	247.691	1	397.471	27.905	64.838	75.220
163.523	1	39	39.755	0	55.902	9.469	26.367	37.867
395.551	1	100	187.258	1	218.828	11.972	78.593	87.475
203.175	1	89	162.186	0	196.590	9.460	41.371	51.670
481.255	1	271	1024.671	1	1138.337	22.416	65.611	71.222

169.726	1	41	59.759	0	78.409	8.431	20.776	37.825
174.049	1	77	80.068	0	174.022	11.938	65.333	69.672
312.143	1	125	431.037	1	528.684	20.313	72.453	87.471
192.998	1	40	101.732	0	157.881	7.473	27.659	34.483
206.534	1	36	133.630	0	218.602	6.774	25.591	32.892
197.046	1	49	70.891	0	88.827	4.290	55.687	58.461
258.321	1	50	168.157	1	214.953	9.214	33.209	37.613
260.365	1	98	311.200	1	394.201	42.412	41.585	49.267
213.738	0	36	133.162	0	175.064	5.755	26.682	31.690
217.342	1	30	71.600	0	93.943	10.108	27.235	32.849
212.759	1	79	115.621	0	207.514	13.762	61.672	69.332
151.200	0	36	91.267	0	121.586	6.448	26.108	32.224
213.170	1	107	110.735	0	178.672	28.942	91.195	97.365
266.298	1	68	175.863	1	223.313	12.903	66.500	72.895
190.425	1	45	63.904	0	103.992	10.199	30.887	40.054
299.558	1	61	115.206	0	152.672	13.358	39.655	45.870
298.743	1	87	136.522	1	228.685	18.134	55.520	62.726
196.100	1	42	45.737	0	49.580	8.149	27.755	43.905
309.039	1	97	485.482	1	619.706	23.351	72.120	94.008
222.110	1	63	176.244	1	208.269	13.284	44.565	52.307
121.500	1	16	15.361	0	26.030	5.021	12.360	15.725
269.801	1	112	145.867	0	229.637	21.557	71.483	80.489
187.806	1	52	58.640	0	105.796	10.492	40.829	46.821
186.000	1	32	71.421	0	101.510	3.182	20.981	24.324
314.970	1	86	500.372	1	588.245	17.326	63.119	75.491
270.276	1	181	627.332	1	764.021	42.432	119.462	146.812
252.347	0	48	164.797	0	224.967	13.439	39.115	45.194
220.000	0	38	82.042	0	117.160	5.990	34.322	39.520
189.418	1	29	51.584	0	91.816	13.266	15.048	17.466
176.366	1	48	76.368	0	113.147	11.662	37.014	43.537
275.768	1	75	100.822	0	181.426	23.042	52.041	74.152
317.083	0	140	388.472	1	505.679	29.474	86.954	97.351
226.190	1	79	80.698	0	197.148	16.119	50.330	58.034
170.322	0	36	53.127	0	68.329	4.878	24.298	31.282
169.919	0	19	32.561	0	60.467	5.082	19.390	23.530
305.391	0	191	595.495	1	703.137	21.194	126.763	140.410
223.600	1	52	255.090	1	320.305	12.561	58.767	76.776
259.061	1	75	123.487	0	163.152	16.438	55.337	70.843
235.687	0	50	47.102	0	69.587	8.002	31.261	35.957
278.379	1	252	1099.930	1	1344.232	21.730	68.838	86.361
206.995	1	40	75.155	0	116.287	10.104	28.928	33.226
183.911	1	39	83.794	0	114.721	5.491	20.275	32.020
192.500	0	35	131.093	0	164.567	8.500	25.863	31.202
297.295	1	99	321.382	0	397.800	17.836	52.156	56.912
332.972	1	103	215.019	1	319.855	41.171	41.874	52.992
177.893	1	22	31.185	0	52.419	9.085	17.520	20.267
277.900	0	58	46.507	1	97.439	16.691	43.405	51.370
249.342	0	47	179.321	1	255.113	7.424	32.118	36.083
539.169	0	248	917.254	1	1225.388	65.696	146.101	173.969
221.558	0	79	228.992	0	298.005	19.908	46.050	57.568
216.611	1	57	85.239	0	154.681	14.783	40.961	44.352
304.305	1	156	949.923	1	1088.206	24.790	73.797	100.057
199.846	0	36	103.914	0	161.424	9.342	29.180	35.249

278.702	1	89	262.667	1	353.379	14.549	81.250	85.054
289.114	0	87	188.427	0	272.778	10.567	75.857	84.273
201.767	1	71	232.928	1	344.778	10.602	58.849	65.564
193.536	1	51	93.059	0	157.691	9.368	42.310	47.097
263.987	1	53	85.508	0	152.527	16.837	37.557	43.579
254.225	1	150	438.392	1	800.677	48.041	83.836	93.107
210.922	1	32	114.200	0	173.806	9.740	21.783	29.424
164.243	0	20	46.455	0	77.837	5.220	9.182	19.529
313.415	1	147	520.674	1	664.662	68.250	98.090	127.676
191.567	1	28	34.529	0	41.398	5.795	25.744	30.071
216.987	1	51	84.883	0	129.076	11.336	34.158	37.161
182.047	1	55	17.248	0	79.992	13.355	34.789	41.081
219.404	1	78	260.567	0	359.070	14.671	58.026	66.837
233.676	0	73	145.548	0	199.144	12.371	50.085	57.643
265.886	1	25	282.156	1	359.427	10.374	43.050	49.165
275.984	0	50	55.034	0	96.096	9.408	36.242	45.861
248.355	1	79	203.637	1	294.873	11.882	71.496	74.314
354.891	1	298	1207.408	1	1460.999	31.201	87.074	110.837
186.674	1	76	105.167	0	130.127	13.785	59.104	68.135
175.230	1	32	102.088	0	153.382	4.557	30.517	32.600