Department of Statistics

COURSE STATS 330

Model answer for Assignment 2, 2007

Ranking Us Colleges by Graduation Rate Performance

A report prepared for US News and World Report by Alan Lee

Shining Light Consultants Inc

Executive summary

A method based on regression analysis is proposed for predicting the graduation rates of US colleges, based on the caliber of the entering student class and the available resources. The colleges are then ranked according as they exceed or fall short of their predicted rate.

Introduction

This report describes a method of measuring the amount of "added value" provided by a college. Added value is defined as the effect of the college's programs and policies on the graduation rate of students after controlling for spending and student aptitude. We measure the difference between a school's six-year graduation rate for the class that entered in 1999 and the rate we predicted for the class. If the actual graduation rate is higher than the predicted rate, the college is enhancing achievement. The actual difference between the actual rate and the predicted rate (the graduation rate performance index) provides a quantitative measure of added value. The colleges are then ranked according to this measure.

Data

The data available for analysis consisted of the following variables, each measured on 173 colleges:

university:	the name of the college,
grad.rate:	the graduation rate (the percentage of the entering class of 2001 that graduated by 2006),
fresh.ret:	the freshman retention rate (percentage of first year students that return),
per20:	percentage of classes with fewer than 20 students,
per50:	percentage of classes with more than 50 students,
SAT75:	75^{th} percentile of the SAT scores of admitted students

(the SAT is a standard college admission test),
top10: percentage of freshmen in top 10% of their high school class,
accept.rate: The percent of applicants accepted for admission,
alumni.giving: The percentage of alumni who contribute money.

Method

The predicted rate was calculated by fitting a regression model to the data, using the variable **grad.rate** as the response variable. Details of the fitting are contained in Appendix 1. The final prediction equation arrived at is

```
Prediction =      0.21679     -1.20055 * fresh.ret

 +  1.632178 * fresh.ret^2  + 0.0001597877 * SAT75

+  0.505659 * alumni.giving  - 0.583171* alumni.giving^2
```

The actual index is the difference between the actual rate and the predicted rate. Note that if all the data were used to fit the regression model, this would be the residual for a particular college.

Results

The index for the top 20 colleges is shown in Table 1. A complete ranked list of all 173 colleges is given in Appendix 2.

Table 1. Top 20 colleges, ranked according to the graduation rate performance index.

University	Index
University of New Hampshire	0.1410
Oregon State University	0.1334
University of Rhode Island	0.1322
Bowling Green State University	0.1236
Pepperdine University	0.1080
Catholic University of America	0.1069
Fordham University	0.1064
Illinois State University	0.0966
Hofstra University	0.0935
Ball State University	0.0905
Miami (OH) University	0.0851
University of California--Santa Barbara	0.0825
Florida Institute of Technology	0.0799
St. Johns University	0.0780
Arizona State University	0.0712
University of Utah	0.0708
Andrews University	0.0707

University of Colorado	0.0707
University of California--Santa Cruz	0.0703
Clemson University	0.0697

A Note of Caution: The rankings given here depend on exactly which model is fitted. Fitting a different model will result in a different ranking. Thus, the exact position in the list should not be interpreted too literally.

Statistical Appendix

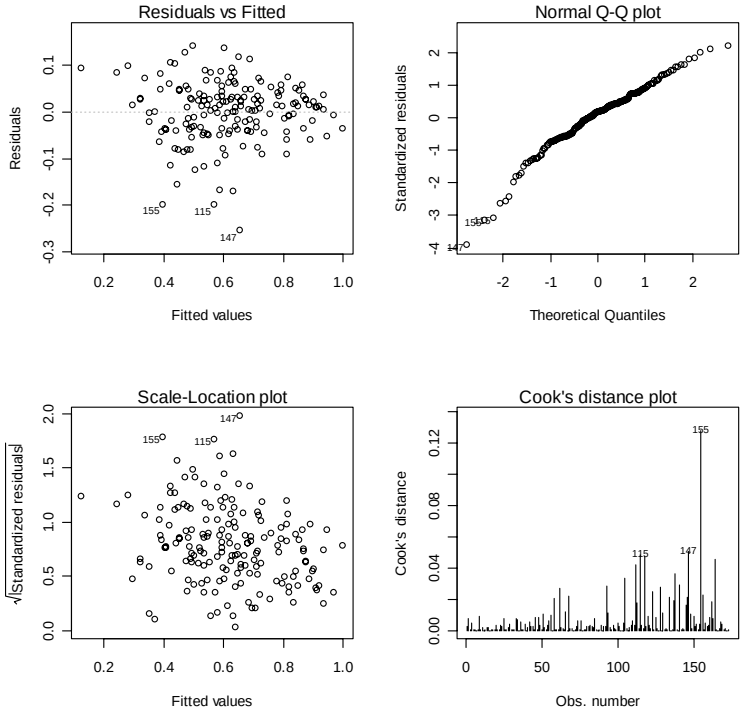
The data were loaded into R from the supplied file **college.csv.** A regression model was fitted, using all the variables, (except the names!). Output is shown below:

```
full.model=lm(grad.rate~., data=colleges.df[,-1])

summary(full.model)
Call:
lm(formula = grad.rate ~ ., data = colleges.df[, -1])

Residuals:
      Min        1Q    Median        3Q       Max
-0.254498 -0.037653  0.009696  0.039694  0.142059

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   -8.499e-01  1.264e-01  -6.726 2.73e-10 ***
fresh.ret      1.281e+00  9.581e-02  13.375  < 2e-16 ***
per20          6.221e-02  6.311e-02   0.986 0.325697
per50          3.744e-02  1.232e-01   0.304 0.761524
SAT75          2.231e-04  9.764e-05   2.285 0.023608 *
top10          4.501e-02  4.101e-02   1.098 0.274018
accept.rate    5.186e-02  4.092e-02   1.267 0.206885
alumni.giving  2.346e-01  6.723e-02   3.490 0.000619 ***
Residual standard error: 0.06582 on 165 degrees of freedom
Multiple R-Squared: 0.8679,     Adjusted R-squared: 0.8623
F-statistic: 154.8 on 7 and 165 DF,  p-value: < 2.2e-16
```



From this, we see that the fit is good, but that there are about 3 influential points. Also, only three explanatory variables (fresh.ret, SAT75 and alumni.giving seem to have much relationship to the response. As a check that these can be dropped, we fitted the model

```
grad.rate~fresh.ret+SAT75+alumni.giving
```

and compared the fit to that of the full model using anova:

```
sub.model = lm(grad.rate~fresh.ret+SAT75+alumni.giving,
                data=colleges.df)

anova(sub.model, full.model)
```

```
Analysis of Variance Table

Model 1: grad.rate ~ fresh.ret + SAT75 + alumni.giving
Model 2: grad.rate ~ fresh.ret + per20 + per50 + SAT75 + top10 + accept.rate +
    alumni.giving
  Res.Df     RSS  Df Sum of Sq      F Pr(>F)
1    169 0.73141
2    165 0.71472   4   0.01669 0.9632 0.4293
```

It seems that the smaller model is OK.

Next, we tried some gam plots, that suggested that quadratic terms in fresh.ret and alumni.giving might be helpful. The model was refitted, and was found to have one large residual (point 147). This was deleted, and the model refitted,

```
final.model.poly=lm(grad.rate~ fresh.ret + I(fresh.ret^2) +SAT75 +
alumni.giving + I(alumni.giving^2),  subset = (1:173)[-147], data=colleges.df)
summary(final.model.poly)
```

resulting in

```
Call:
lm(formula = grad.rate ~ fresh.ret + I(fresh.ret^2) + SAT75 +
    alumni.giving + I(alumni.giving^2), data = colleges.df,
    subset = (1:173)[-147])

Residuals:
      Min        1Q    Median        3Q       Max
-0.201722 -0.034101  0.004725  0.037783  0.140959

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.168e-01  3.071e-01   0.706 0.481244
fresh.ret          -1.201e+00  7.146e-01  -1.680 0.094821 .
I(fresh.ret^2)      1.632e+00  4.651e-01   3.509 0.000579 ***
SAT75               1.598e-04  7.694e-05   2.077 0.039361 *
alumni.giving       5.057e-01  1.476e-01   3.425 0.000774 ***
I(alumni.giving^2) -5.832e-01  2.595e-01  -2.247 0.025946 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06059 on 166 degrees of freedom
Multiple R-Squared: 0.8863,     Adjusted R-squared: 0.8829
F-statistic: 258.9 on 5 and 166 DF,  p-value: < 2.2e-16
```

This model was used to make the predictions.

Note that we used a data set with 172 data points to fit the model, but want to calculate residuals for 173. We do this using the code

```
grad.index = college.df$grad.rate – predict(final.model, newdata=college.df)
```

We then print out the full ordered list using the code

```
ordered.list=data.frame(University=colleges.df$University,
       Index = grad.index)[order(-grad.index),]
```

Note the minus sign – this is to sort from highest to lowest.

What would have happened if we had fitted a different model, say the full model we fitted first, with no diagnostic checking? The table below shows the ranks of the universities (i.e. 1=best, 2=next best) for the final model and the initial model. There is a reasonable amount of agreement, but this shows that we should be a bit cautious about the exact position in the ranking.

	Final Model	Initial Model
University of New Hampshire	1	2
Oregon State University	2	12
University of Rhode Island	3	1
Bowling Green State University	4	3
Pepperdine University	5	10
Catholic University of America	6	15
Fordham University	7	4
Illinois State University	8	6
Hofstra University	9	11
Ball State University	10	7
Miami (OH) University	11	5
University of California--Santa Barbara	12	36
Florida Institute of Technology	13	38
St. Johns University	14	14
Arizona State University	15	26
University of Utah	16	8
Andrews University	17	16
University of Colorado	18	30
University of California--Santa Cruz	19	63
Clemson University	20	17

This was produced with the code

```
grad.index2 = colleges.df$grad.rate - predict(full.model, newdata=colleges.df)
ranks = cbind(rank(-grad.index), rank(-grad.index2))
dimnames(ranks) = list(colleges.df$University,
                 c("Final Model","Initial Model"))
ranks[order(-grad.index), ][1:20,]
```

Appendix 2: Complete ordered list based on final model

	University	Index
94	University of New Hampshire	0.1410
118	Oregon State University	0.1334
134	University of Rhode Island	0.1322
105	Bowling Green State University	0.1236
66	Pepperdine University	0.1080
51	Catholic University of America	0.1069
58	Fordham University	0.1064
148	Illinois State University	0.0966
110	Hofstra University	0.0935
146	Ball State University	0.0905
62	Miami (OH) University	0.0851
46	University of California--Santa Barbara	0.0825
109	Florida Institute of Technology	0.0799
120	St. Johns University	0.0780
104	Arizona State University	0.0712
141	University of Utah	0.0708
145	Andrews University	0.0707
82	University of Colorado	0.0707
81	University of California--Santa Cruz	0.0703
54	Clemson University	0.0697
49	Auburn University	0.0695
89	University of Massachusetts	0.0687
100	University of Vermont	0.0675
48	American University	0.0660
168	University of Northern Colorado	0.0577
80	University of California--Riverside	0.0560
42	University of Illinois	0.0525
84	University of Denver	0.0483
79	University of Arizona	0.0470
74	State University of New York--Binghamton	0.0454
152	Montana State University	0.0447
111	Indiana University of Pennsylvania	0.0443
70	Southern Methodist University	0.0426
39	University of California--Irvine	0.0424
65	Ohio State University	0.0422
33	College of William and Mary	0.0416
83	University of Delaware	0.0415
167	University of North Carolina--Greensboro	0.0414
156	Portland State University	0.0411
78	Texas Christian University	0.0405
1	Harvard University	0.0399
139	University of the Pacific	0.0398
154	Northern Illinois University	0.0386
144	West Virginia University	0.0375
160	University of Central Florida	0.0374
31	Brandeis University	0.0372
44	University of California--Davis	0.0370
136	University of South Carolina	0.0365
119	Seton Hall University	0.0353
98	University of Tennessee	0.0322
53	Clark University	0.0319
38	Tulane University	0.0319
140	University of Tulsa	0.0295
96	University of Pittsburgh	0.0294

	University	Index
2	Princeton University	0.0280
61	Loyola University--Chicago	0.0275
86	University of Hawaii	0.0274
12	Dartmouth College	0.0256
173	Western Michigan University	0.0244
14	Johns Hopkins University	0.0234
71	Stevens Institute of Technology	0.0229
106	Colorado State University	0.0207
28	University of Michigan	0.0207
129	University of Mississippi	0.0202
101	Virginia Tech	0.0181
73	State University of New York--Albany	0.0180
60	Iowa State University	0.0157
32	University of California--San Diego	0.0154
36	Boston College	0.0152
6	Cornell University	0.0151
20	Georgetown University	0.0147
103	Worcester Polytechnic Institute	0.0132
21	Vanderbilt University	0.0131
87	University of Kansas	0.0120
165	University of Missouri--St. Louis	0.0115
107	DePaul University	0.0099
43	Pennsylvania State University	0.0099
19	University of Notre Dame	0.0097
23	University of Virginia	0.0091
8	University of Pennsylvania	0.0081
35	New York University	0.0078
30	Wake Forest University	0.0074
157	Southern Illinois University	0.0074
57	Florida State University	0.0057
90	University of Miami (FL)	0.0056
10	Brown University	0.0055
149	Indiana State University	0.0039
16	Emory University	0.0032
24	University of North Carolina--Chapel Hill	0.0030
102	Washington State University	0.0026
59	Indiana University	0.0015
3	Yale University	0.0007
13	Northwestern University	-0.0009
135	University of San Francisco	-0.0011
69	Rutgers--New Brunswick	-0.0020
11	Columbia University	-0.0026
37	Lehigh University	-0.0035
77	Texas A&M University	-0.0037
127	University of Maine	-0.0042
124	University of Idaho	-0.0046
75	State University of New York--Buffalo	-0.0073
172	University of Wisconsin--Milwaukee	-0.0082
67	Purdue University	-0.0091
17	Washington University (St. Louis)	-0.0106
40	University of Wisconsin	-0.0112
122	University of Alabama--Huntsville	-0.0113
7	Duke University	-0.0117
26	Tufts University	-0.0138
142	University of Wyoming	-0.0146
88	University of Maryland	-0.0171

	University	Index
91	University of Minnesota	-0.0173
170	University of South Florida	-0.0174
5	Stanford University	-0.0186
153	Northern Arizona University	-0.0200
166	University of Montana	-0.0207
117	Oklahoma State University	-0.0211
99	University of Texas	-0.0213
50	Brigham Young University	-0.0215
22	University of California--Berkeley	-0.0227
18	Rice University	-0.0231
15	University of Chicago	-0.0237
9	California Institute of Technology	-0.0238
131	University of Nebraska	-0.0256
112	Indiana/Purdue--Indianapolis	-0.0257
63	Michigan Technological University	-0.0266
72	St. Louis University	-0.0281
143	Virginia Commonwealth University	-0.0314
171	University of Toledo	-0.0324
114	Mississippi State University	-0.0340
45	Syracuse University	-0.0345
132	University of New Mexico	-0.0355
133	University of North Dakota	-0.0369
52	Clarkson University	-0.0393
27	University of California--Los Angeles	-0.0393
92	University of Missouri	-0.0398
159	University of Akron	-0.0398
76	State University of New York--Stony Brook	-0.0409
121	University of Alabama	-0.0415
169	University of North Texas	-0.0417
95	University of Oklahoma	-0.0431
151	Middle Tennessee State University	-0.0444
108	Drexel University	-0.0447
126	University of Kentucky	-0.0478
29	University of Rochester	-0.0479
93	University of Missouri--Rolla	-0.0480
47	George Washington University	-0.0487
34	Case Western University	-0.0519
130	University of Missouri--Kansas City	-0.0555
4	Massachusetts Inst. Of Technology	-0.0568
64	North Carolina State University	-0.0606
55	Colorado School of Mines	-0.0637
41	University of Southern California	-0.0639
150	Louisiana Tech University	-0.0644
97	University of San Diego	-0.0655
158	Texas Tech University	-0.0669
125	University of Illinois--Chicago	-0.0702
163	University of Houston	-0.0758
56	Duquesne University	-0.0822
116	Northeastern University	-0.0837
137	University of Southern Mississippi	-0.0847
25	Carnegie Mellon University	-0.0899
138	University of Texas--Dallas	-0.1028
85	University of Florida	-0.1055
161	University of Colorado--Denver	-0.1102
113	Louisiana State University	-0.1149

	University	Index
162	University of Detroit--Mercy	-0.1179
68	Rutgers--Newark	-0.1198
155	Nova Southeastern University	-0.1590
123	University of Arkansas	-0.1629
128	University of Maryland--Baltimore Co.	-0.1680
164	University of Louisville	-0.1748
115	New Jersey Institute of Technology	-0.2017
147	Florida International University	-0.2662