Forecasting College Enrollment Using SAS®/ETS Procedures

J. Richard Earley, American College for the Applied Arts, Los Angeles, California

INTRODUCTION

This paper investigates use of the SAS System in analyzing college enrollment activity to forecast enrollment statistics in relation to acceptances and applications. Consideration was also given to tuition, student-teacher ratios, and to public versus private institutions. Data was compiled for eighty-five California colleges and one-hundred and eleven colleges in New York, using Lovejoy's College Guide, copyright 1992, and including every college in Califor/nia and New York listing statistics on freshman enrollments. Data was also compiled on nine Ivy League colleges. The SAS System was then used as the primary tool for developing and estimating statistical models. In particular, PROC RSQUARE, PROC REG, PROC CORR, PROC ARIMA and PROC STEPWISE were used in addition to the usual descriptive statistics provided by PROC MEANS, PROC CHART and PROC UNIVARIATE.

An econometric model was developed using relevant variables such as student-teacher ratios, applications to enrollment ratios, applications to acceptances ratios, and enrollment to acceptances ratios as the regressors. The paper then presents and analyzes summary measures such as R^2 and t - values to assess statistical significance.

THEORY

Visual inspection of the data before analysis indicated that less than half of students who are accepted by colleges actually enroll, leading to question what relationships might exist between student enrollments and variables such as student-teacher ratios, tuition (including fees), and the colleges' acceptances of students. Moreover, did significances appear between California and New York, two of our largest upper educational systems, that might lead to possible directions in helping colleges forecasting enrollments or in establishing more effective procedures in the enrollment process? For example, do differences appear between the colleges with the highest ratios of enrollments to acceptances and those with the lowest? Or do colleges that accept the highest percentages of students differ on these variables from colleges accepting the fewest students? If so, what data might help administrators enhance or streamline the enrollment process?

What might California's college education look like compared to New York's in average and minimum and maximum student-teacher ratios? What of educational costs averages and variances by state, and any major differences between public and private colleges in the states, such as in student-teacher ratios or in higher percentages of those students who were accepted actually enrolling?

Finally, further insights on enrollment practices were sought by comparing the results from California and New York against a selected group of nine ivy league colleges: Brown, Colgate, Columbia, Cornell, Dartmouth, Harvard, Penn, Princeton, and Yale.

EMPIRICAL RESULTS

California Data

Simple Statistics

Variable	N	Mean	Std Dev	Sum
TYPE	85	0.68235	0.46832	58.00000
TUITION	85	7562	5032	642733
SFRAT	83	14.68675	5.07006	1219
APPLY	82	3395	5189	278388
ACCEPT	82	2162	2982	177256
ENROLL	83	802.93976	990	66644
APPTOACC	81	1.53047	0.55219	123.96824
ACCTOAPP	81	0.70776	0.16817	57.32825
ETOACCEP	80	0.48396	0.18543	38.71542
ETOAPP	80	0.35468	0.18361	28.37433
NUMBER	85	43.00000	24.68130	3655

Simple Statistics

Variable	Minimum	Maximum
TYPE	0	1.00000
TUITION	549.00000	17380
SFRAT	3.00000	35.00000
APPLY	36.00000	22610
ACCEPT	33.00000	11039
ENROLL	13.00000	3636
APPTOACC	1.00000	4.54545
ACCTOAPP	0.22000	1.00000
ETOACCEP	0.20249	1.00000
ETOAPP	0.12355	0.90909
NUMBER	1.00000	85.00000

Regression Analysis

Econometric Model for Applications:

The following three regression models were estimated for each of the data sets:

$$Apply = \beta_0 + \beta_1 APPTOACC + \beta_2 SFRAT + \beta_3 TUITION + \beta_4 TYPE + u_t$$

$$Enroll = \beta_0 + \beta_1 ACCTOAPP + \beta_2 SFRAT + \beta_3 TUITION + \beta_4 TYPE + u_t$$

$$ETOACCEP = \beta_0 + \beta_1 TYPE + \beta_2 TUITION + \beta_3 SFRAT + u_t$$

Model: MODEL1
Dependent Variable: APPLY

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	1347639339.8	336909834.94	30.437	0.0001
Error	73	808052928.20	11069218.195		
C Total	77	2155692268.0			

Root MSE	3327.04346	R-square	0.6252
Dep Mean	3391.11538	Adj R-sq	0.6046
C.V.	98.11059		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	249.247122	2416.7177624	0.103	0.9181
APPTOACC	1	3732.125149	806.59515861	4.627	0.0001
SFRAT	1	165.633134	89.33537639	1.854	0.0678
TUITION	1	0.060828	0.15602957	0.390	0.6978
TYPE	1	-7570.815832	1678.3771320	-4.511	0.0001

Durbin-Watson D		1.234
(For Number of Obs.)		78
1st Order Autocorrelation		0.377

Model: MODEL2
Dependent Variable: ENROLL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	49146904.165	12286726.041	34.698	0.0001
Error	73	25849369.950	354100.95822		
C Total	77	74996274.115			

Root MSE	595.06383	R-square	0.6553
Dep Mean	771.19231	Adj R-sq	0.6364
C.V.	77.16154		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	1955.842365	421.99351576	4.635	0.0001
ACCTOAPP	1	-761.197725	504.34373279	-1.513	0.1345
SFRAT	1	29.542432	16.03960608	1.842	0.0696
TUITION	1	0.020854	0.02904994	0.718	0.4751
TYPE	1	-1723.173539	319.85627196	-5.387	0.0001

Durbin-Watson D 2.098
(For Number of Obs.) 78
1st Order Autocorrelation -0.069

Model: MODEL3
Dependent Variable: ETOACCEP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	1.19529	0.39843	19.781	0.0001
Error	74	1.49053	0.02014		
C Total	77	2.68582			

Root MSE	0.14192	R-square	0.4450
Dep Mean	0.48575	Adj R-sq	0.4225
C.V.	29.21747		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.599249	0.07609738	7.875	0.0001
TYPE	1	0.455877	0.06361008	7.167	0.0001
TUITION	1	-0.000063488	0.00000604	-7.199	0.0001
SFRAT	1	-0.006593	0.00365483	-1.804	0.0753

Durbin-Watson D 1.712
(For Number of Obs.) 78
1st Order Autocorrelation 0.143

New York Data

Simple Statistics

Variable	N	Mean	Std Dev	Sum
TYPE	111	0.31532	0.46675	35.00000
TUITION	111	8223	4846	912767
SFRAT	111	14.54054	4.65888	1614
APPLY	109	3614	3846	393896
ACCEPT	107	1980	1853	211895
ENROLL	111	767.39640	740.34805	85181
APPTOACC	105	1.77176	1.01753	186.03480
ACCTOAPP	105	0.65296	0.18486	68.56088
ETOACCEP	107	0.41144	0.14838	44.02428
ETOAPP	109	0.26874	0.12634	29.29217
NUMBER	111	56.00000	32.18695	6216

Simple Statistics

Variable	Minimum	Maximum
TYPE	0	1.00000
TUITION	0	17312
SFRAT	4.00000	30.00000
APPLY	165.00000	20326
ACCEPT	88.00000	8428
ENROLL	40.00000	3410
APPTOACC	1.03186	7.40447
ACCTOAPP	0.13505	0.96912
ETOACCEP	0.21809	0.90643
ETOAPP	0.08935	0.74375
NUMBER	1.00000	111.00000

Model: MODEL1
Dependent Variable: APPLY

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	433492093.20	108373023.30	9.514	0.0001
Error	100	1139060786.2	11390607.862		
C Total	104	1572552879.4			

Root MSE	3374.89746	R-square	0.2757
Dep Mean	3578.71429	Adj R-sq	0.2467
C.V.	94.30754		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-5528.299951	2368.4671760	-2.334	0.0216
APPTOACC	1	1613.295387	379.12747419	4.255	0.0001
SFRAT	1	282.132664	99.34543811	2.840	0.0055
TUITION	1	0.171576	0.09686413	1.771	0.0796
TYPE	1	2555.202915	903.61694856	2.828	0.0057

Durbin-Watson D 1.625
(For Number of Obs.) 105
1st Order Autocorrelation 0.187

Model: MODEL2
Dependent Variable: ENROLL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	7440043.8922	1860010.9731	4.171	0.0036
Error	100	44598804.641	445988.04641		
C Total	104	52038848.533			

Root MSE	667.82336	R-square	0.1430
Dep Mean	739.93333	Adj R-sq	0.1087
C.V.	90.25453		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	375.804144	386.86711303	0.971	0.3337
ACCTOAPP	1	-713.486058	395.53502485	-1.804	0.0743
SFRAT	1	52.174408	18.48696182	2.822	0.0058
TUITION	1	0.004898	0.01836220	0.267	0.7902
TYPE	1	144.328882	185.27301990	0.779	0.4378

Durbin-Watson D 1.719
(For Number of Obs.) 105
1st Order Autocorrelation 0.128

Model: MODEL3
Dependent Variable: ETOACCEP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	0.45121	0.15040	8.181	0.0001
Error	101	1.85683	0.01838		
C Total	104	2.30804			

Root MSE	0.13559	R-square	0.1955
Dep Mean	0.41075	Adj R-sq	0.1716
C.V.	33.00997		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.737096	0.07136891	10.328	0.0001
TYPE	1	-0.056939	0.03590185	-1.586	0.1159
TUITION	1	-0.000017230	0.00000371	-4.649	0.0001
SFRAT	1	-0.011499	0.00349085	-3.294	0.0014

Durbin-Watson D 1.408
(For Number of Obs.) 105
1st Order Autocorrelation 0.290

Ivy League Data

Simple Statistics

Variable	N	Mean	Std Dev	Sum
TYPE	9	1.00000	0	9.00000
TUITION	9	16506	719.45031	148554
SFRAT	8	9.12500	2.10017	73.00000
APPLY	9	10809	4480	97285
ACCEPT	9	2949	1477	26545
ENROLL	9	1498	743.15147	13483
APPTOACC	9	3.92170	1.37189	35.29532
ACCTOAPP	9	0.28727	0.10882	2.58542
ETOACCEP	9	0.51303	0.11230	4.61731
ETOAPP	9	0.13975	0.03981	1.25773
NUMBER	9	5.00000	2.73861	45.00000

Simple Statistics

Variable	Minimum	Maximum
TYPE	1.00000	1.00000
TUITION	15858	18180
SFRAT	6.00000	12.00000
APPLY	5193	20326
ACCEPT	1960	6238
ENROLL	718.00000	3042
APPTOACC	2.12666	5.88192
ACCTOAPP	0.17001	0.47022
ETOACCEP	0.31340	0.74387
ETOAPP	0.09350	0.23649
NUMBER	1.00000	9.00000

Model: MODEL1
Dependent Variable: APPLY

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	24512014.941	8170671.6470	0.240	0.8645
Error	4	136049410.93	34012352.734		
C Total	7	160561425.87			

Root MSE	5832.01104	R-square	0.1527
Dep Mean	10811.37500	Adj R-sq	-0.4828
C.V.	53.94329		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	3002.172561	69451.904554	0.043	0.9676
APPTOACC	1	1582.106931	2845.6205501	0.556	0.6079
SFRAT	1	573.253542	1251.5799278	0.458	0.6707
TUITION	1	-0.212487	4.81326777	-0.044	0.9669

Durbin-Watson D 2.819
(For Number of Obs.) 8
1st Order Autocorrelation -0.411

Model: MODEL2
Dependent Variable: ENROLL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	243512.27696	81170.75899	0.079	0.9682
Error	4	4125956.5980	1031489.1495		
C Total	7	4369468.8750			

Root MSE	1015.62254	R-square	0.0557
Dep Mean	1524.12500	Adj R-sq	-0.6525
C.V.	66.63643		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-2347.931144	12776.774382	-0.184	0.8631
ACCTOAPP	1	1338.037262	5030.5718443	0.266	0.8034
SFRAT	1	60.977327	197.28762614	0.309	0.7727
TUITION	1	0.177130	0.70975715	0.250	0.8152

Durbin-Watson D		2.435
(For Number of Obs.)		8
1st Order Autocorrelation		-0.218

Model: MODEL3
Dependent Variable: ETOACCEP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.07461	0.03731	7.402	0.0320
Error	5	0.02520	0.00504		
C Total	7	0.09981			

Root MSE	0.07099	R-square	0.7475
Dep Mean	0.50918	Adj R-sq	0.6468
C.V.	13.94288		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-1.485571	0.63533958	-2.338	0.0665
TUITION	1	0.000126	0.00003609	3.496	0.0173
SFRAT	1	-0.009389	0.01316708	-0.713	0.5077

Durbin-Watson D		1.187
(For Number of Obs.)		8
1st Order Autocorrelation		0.364

Summary of Regressions

Model 1: Cal Applications R2 = 0.6252 Variables: Apptoacc = 4.627* Sfrat = 1.854 Tuition = 0.390 Type = -4.511	New York R2 = 0.2757 Variables: Apptoacc = 4.255 Sfrat = 2.840 Tuition = 1.771 Type = 2.828	Ivy League R2 = 0.1527 Variables: Apptoacc = 3.556 Sfrat = 0.458 Tuition =-0.044
Model 2: Enrollments R2 = 0.6553 Variables: Acctoapp = -1.513 Sfrat = 1.842 Tuition = 0.718 Type = -5.387	R2 = 0.1430 Variables: Acctoapp = -1.804 Sfrat = 2.822 Tuition = 0.267 Type = 0.779	R2 = 0.0557 Variables: Acctoapp =0.266 Sfrat = 0.309 Tuition =0.250
Model 3: Enrollments to Acceptances R2 = 0.4450 Variables: Sfrat = -1.804 Tuition = -7.199 Type = 7.167	R2 = 0.1955 Variables: Sfrat = -3.294 Tuition = -4.649 Type = -1.586	R2 = 0.7475 Variables: Sfrat = -0.713 Tuition = 3.496

Comparison of Groups

	Cal	New York	Ivy
Applied	278,390	393,898	97,258
Accepted	177,256	211,895	26,545
Enrolled	66,666	85,181	13,483
%ofApps Accepted	63.6	53.9	27.3
%ofAccepts Enrolled	37.6	40.2	50.8
%of Applying Enrolled	23.9	21.9	13.8
SFRatio	14.68	14.54	9.1
Tuition	$7,562	$8,223	$16,506
Public(0)vs Private (1)	0.68	0.31	1.00

Conclusion

Analysis of the PROCREG output leads to the belief that significances exist in this analysis of college enrollments. For California colleges the coefficient of determination, R-square, was equal to 0.6252, indicating that approximately 63% of the Applications variable was accounted for or explained by the variables, of which the Apptoacc and Type were statistically significant as judged by the t-values. For the Enrollment model, the R-square was equal to 0.6553, with the Type of institution significant. The Etoaccep model had an R-square of 0.4450, with the Type and Tuition variables significant.

For New York the Apply model had an R-square of 0.2757, with the variables Apptoacc, Sfratio and Type having significant t-values. R-square for the Enroll model was 0.1430, with the Sfratio significant. The Etoaccep model R-square was 0.1955, with tuition and Sfratio significant.

For the Ivy League, the Apply model R-square was 0.1527, with no significant t-values. The Enroll R-square was 0.0557 with no significances. And the Etoaccep variable had an R-square of 0.7475, indicating that approximately 75% of the variation was explained. The Tuition variable was significant.

Much information was revealed about college education in New York and California. While student-teacher ratios were quite close between the states, with California's average 14.69 (minimum 3, maximum 35) versus New York's mean student-teacher ratio of 14.54 (minimum 4, maximum 30). Although maximum tuitions - including fees - were close in each state (California $17,380, New York $17,312), California colleges were on average less expensive, charging a mean tuition of $7,562, almost seven hundred dollars lower than New York's average of $8,223.

As to indications of institutional size differences, California averaged 803 freshman enrollments per college (minimum 13, maximum 3636) versus New York's per college average of 767 (minimum 40, maximum 3410). These freshmen were chosen from a minimum of 36 and a maximum of 22,610 applying in California, and a mimimum of 165 applicants and a maximum 20,326 in New York. Total applicants in the study numbered 278,390 for California and 393,898 for New York. Total enrollments were 66,666 (23.9% of those applying) for California and 85181 (21.9% of applicants) for New York.

There was little difference between minimum and maximum enrollments as percentages of acceptances, with California's minimum (20% of those accepted actually enrolled) and maximum (100% of those accepted enrolled) and New York's minimum (21.8% of those accepted enrolled) and maximum (90.6% of those accepted enrolled).

In the overall enrollment process, the mean averages for New York were: 3,614 applied, 1,980 were accepted, and 767 actually enrolled. The means in California were: 3,395 applied, 2,162 were accepted, and 803 actually enrolled. The total enrollment figures show California with a lower actual freshman enrollment ratio of 37.6% of those accepted versus New York's 40.2%, and 23.9% enrollment of all applicants versus New York's 21.9% enrollment of all applying. California also had a lower applicant/acceptance ratio of only 1.53 versus New York's 1.77 and the Ivy League's 3.92.

REFERENCES

Gujarati, Damodar. Basic Econometrics. Second edition, McGraw-Hill, 1988.

SAS Institute Inc. SAS System for Elementary Statistical Analysis, 1987.

SAS Institute Inc., SAS/ETS User's Guide. 1988 Version 6, First Edition, Cary, NC: SAS Institute Inc., 1988.

SAS Institute Inc., SAS/ETS Software: Applications Guide 1, Version 6, First Edition, 1991.

SAS Institute Inc., SAS/STAT User's Guide, Vols I and II, 1990 Version 6, Fourth Edition, Cary, N.C.: SAS Institute Inc.