STAT 223

Laboratory Activity #5 by Akbota Serikkyzy

Introduction

It contains data on 777 universities and colleges in the US. The dataset includes the following variables, which we will use to analyze their relationships:

Private: Public/private indicator
Apps: Number of applications received
Accept: Number of applicants accepted
Enroll: Number of new students enrolled
Top10perc: New students from top 10% of high school class
Top25perc: New students from top 25% of high school class
FUndergrad: Number of full-time undergraduates
PUndergrad: Number of part-time undergraduates
Outstate: Out-of-state tuition
RoomBoard: Room and board costs
Books: Estimated book costs
Personal: Estimated personal spending
PhD: Percent of faculty with Ph.D.s
Terminal: Percent of faculty with terminal degree
SFRatio: Student/faculty ratio
percalumni: Percent of alumni who donate
Expend: Instructional expenditure per student
GradRate: Graduation rate

Logistic Regressions on Elite

I will be performing logistic regression on elite with the variables Terminal and Expend to evaluate how the model performs and which variable is more effective. I chose Terminal and Expend because I assumed that top students are likely to be accepted into many institutions, so they would prefer universities with high terminal values and higher expenditures, as they aim to receive the best possible education.

Terminal

$$P(\text{EliteNum} = 1 \mid \text{Terminal}) = \frac{1}{1 + e^{-(-19.0355 + 0.1896 \times \text{Terminal})}}$$

$$\text{Decision Boundary} = -\frac{\beta_0}{\beta_1} = -\frac{-19.0355}{0.1896} \approx 100.37$$

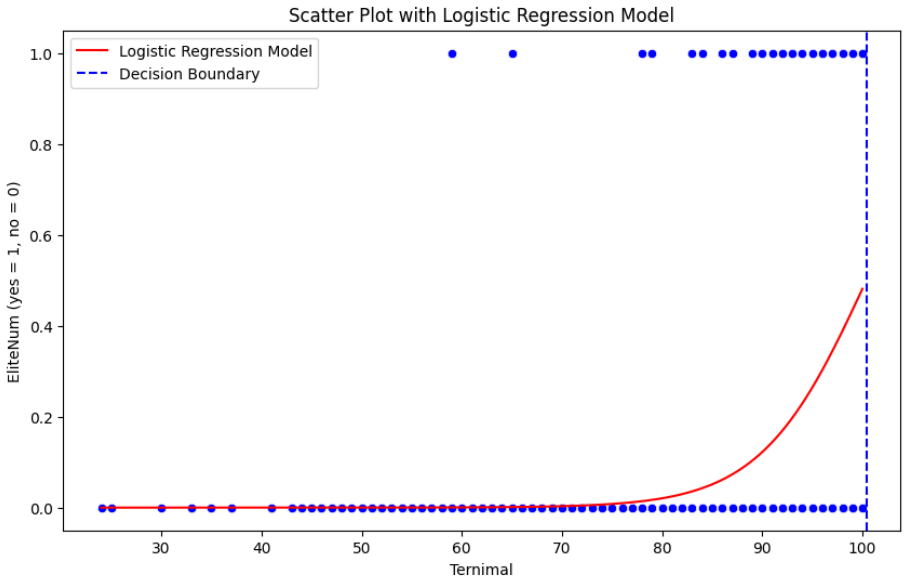

Figure 1: Scatter Plot with Logistic Regression Model

As can be seen in the formula, being elite has a positive association with terminal, with a coefficient of 0.1896. This suggests that the higher the terminal value, the more likely it is to be classified as elite. The decision boundary is 100.37, as also shown in figure 1. If the terminal value is less than 100.37, the model predicts the case as not elite; if it is higher than 100.37, it is predicted to be elite. Additionally, as can be seen in figure 1, most of the data points are not elite, with only a few being classified as elite, indicating an imbalance in the data.

Expend

$$P(\text{EliteNum} = 1 \mid \text{Expend}) = \frac{1}{1 + e^{-(-6.0704 + 0.0003319 \times \text{Expend})}}$$

$$\text{Decision Boundary} = -\frac{\beta_0}{\beta_1} = -\frac{-6.0704}{0.0003319} \approx 18293.66$$

As shown in the formula, Expend has a positive association with being elite, with a coefficient of 0.0003319. This indicates that universities spending more per student are more likely to be classified as elite. The decision boundary is approximately 18 293.66, meaning that universities with expenditures above this threshold are predicted to be elite, while those below are not. As can be seen in Figure 2, the logistic regression model follows an S-shaped curve, indicating that the probability that a university is classified as elite increases gradually with expenditure.

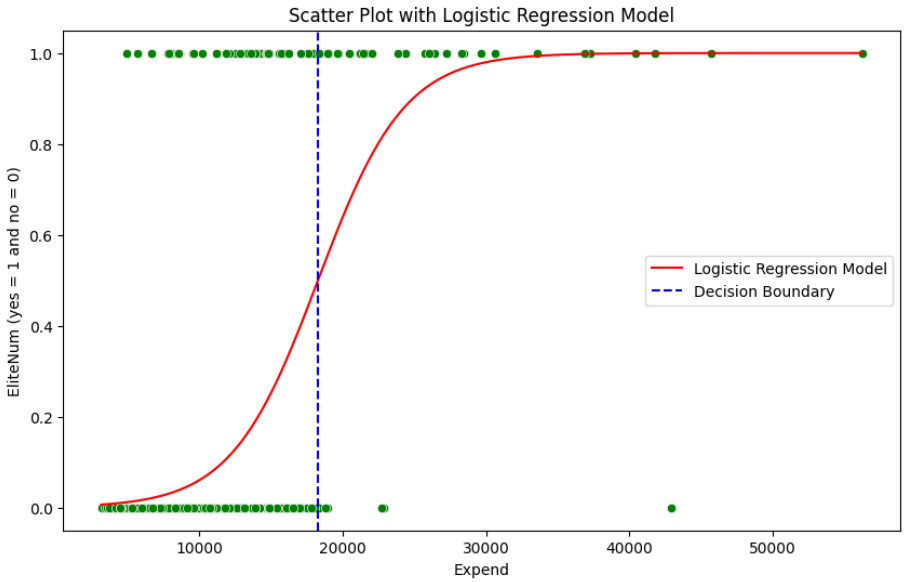

Figure 2: Scatter Plot with Logistic Regression Model

percalumni

I decided to choose Expend, Top10perc, Private, Terminal, and GradRate because I thought they might affect the percentage of alumni who donate. My reasoning is that if a university spends a good amount of money on students, those students might feel more willing to give back. Also, high spending likely means students had a comfortable experience during their time at the university, which could lead to higher satisfaction. I believe that a sense of satisfaction can lead to donations. I also assumed that Top10perc, the percentage of students in the top 10% of their high school class, could be another factor. If they performed well in high school, they might also succeed in college and land good jobs, making them more able and likely to donate. I chose Private because, in a research article I read about alumni donations, it was mentioned that private schools tend to receive more donations. I included Terminal because I thought that having more highly qualified professors might improve student satisfaction, which could lead to higher alumni donations later on. Lastly, I picked GradRate because I think the graduation rate reflects how well a university retains its students, which again can be a sign of satisfaction.

Expend and percalumni

$$\text{percalumni} = 13.1680 + 0.0010 \times \text{Expend}$$

As shown in table 1, Expend has a positive and statistically significant relationship with alumni donation percentage with a coefficient of 0.0010 and a p-value of 0.000. This means that as a university spends more per student, the percentage of alumni who donate tends to increase. In the table 2, the R-squared value is 0.174, which means that about 17.4% of the variation in alumni

Variable	Coefficient	Standard Error	p-value
Intercept	13.1680	0.850	0.000
Expend	0.0010	7.75e-05	0.000

Table 1: Regression Results for Alumni Donation Percentage

Statistic	Value
R-squared	0.174
Adjusted R-squared	0.173
F-statistic	163.8
Prob (F-statistic)	3.67e-34
AIC	5971.0
BIC	5980.0
Durbin-Watson	1.866

Table 2: Model Summary Statistics for OLS Regression

donations can be explained by Expend alone. Although this is not super high, it still shows that Expend has some meaningful effect on alumni giving behavior.

Top10Perc and percalumni

Variable	Coefficient	Standard Error	p-value
Intercept	13.9261	0.735	0.000
Top10perc	0.3200	0.022	0.000

Table 3: Regression Results for Alumni Donation Percentage using Top 10% of High School Class

Statistic	Value
R-squared	0.207
Adjusted R-squared	0.206
F-statistic	202.9
Prob (F-statistic)	4.63e-41
AIC	5939.0
BIC	5948.0
Durbin-Watson	1.782

Table 4: Model Summary Statistics for OLS Regression

$$\text{percalumni} = 13.9261 + 0.3200 \times \text{Top10perc}$$

As shown in table 3, Top10perc has a positive and statistically significant relationship with the percentage of alumni who donate, with a coefficient of 0.3200 and a p-value of 0.000. This suggests

that as the percentage of students in the top 10% of their high school class increases, the percentage of alumni donations also tends to increase. In table 4, the R-squared value is 0.207, which means that about 20.7% of the variation in alumni giving can be explained by Top10perc alone. Although it doesn't capture all the variation, it indicates that student academic quality is a meaningful predictor of alumni donation behavior.

Private and percalumni

Variable	Coefficient	Standard Error	p-value
Intercept	14.3585	0.775	0.000
Private[T.Yes]	11.5318	0.909	0.000

Table 5: Regression Results for Alumni Donation Percentage by Private vs. Public Institutions

Statistic	Value
R-squared	0.172
Adjusted R-squared	0.171
F-statistic	161.0
Prob (F-statistic)	1.16e-33
AIC	5973.0
BIC	5982.0
Durbin-Watson	1.960

Table 6: Model Summary Statistics for OLS Regression

$$\text{percalumni} = 14.3585 + 11.5318 \times \text{Private}$$

As shown in table 5, being a private institution has a strong and statistically significant association with the percentage of alumni donation. The coefficient for `Private[T.Yes]` is 11.5318 with a p-value of 0.000, which means that private universities tend to have about 11.5 percentage points higher alumni donation rates compared to public ones. In table 6, the R-squared value is 0.172, indicating that around 17.2% of the variation in alumni giving can be explained by whether a school is private. Although this model does not account for everything, it still highlights the importance of institutional type in understanding donation patterns.

Terminal and percalumni

$$\text{percalumni} = 4.8233 + 0.2248 \times \text{Terminal}$$

As shown in Table 7, the terminal variable, which represents the percentage of faculty with terminal degrees, has a positive and statistically significant relationship with the percentage of alumni donation. The coefficient is 0.2248 with a p-value of 0.000, indicating that as the percentage of

Variable	Coefficient	Standard Error	p-value
Intercept	4.8233	2.361	0.041
Terminal	0.2248	0.029	0.000

Table 7: Regression Results for Alumni Donation Percentage using Terminal Degree Percentage

Statistic	Value
R-squared	0.071
Adjusted R-squared	0.070
F-statistic	59.55
Prob (F-statistic)	3.67e-14
AIC	6062.0
BIC	6071.0
Durbin-Watson	1.696

Table 8: Model Summary Statistics for OLS Regression

highly qualified faculty increases, the percentage of alumni who donate also tends to rise. According to table 8, the R-squared value is 0.071, meaning that around 7.1% of the variation in alumni donations can be explained by faculty qualifications alone.

GradRate and percalumni

Variable	Coefficient	Standard Error	p-value
Intercept	-0.4385	1.528	0.774
GradRate	0.3541	0.023	0.000

Table 9: Regression Results for Alumni Donation Percentage using Graduation Rate

Statistic	Value
R-squared	0.241
Adjusted R-squared	0.240
F-statistic	246.1
Prob (F-statistic)	2.30e-48
AIC	5905.0
BIC	5915.0
Durbin-Watson	1.956

Table 10: Model Summary Statistics for OLS Regression

$$\text{percalumni} = -0.4385 + 0.3541 \times \text{GradRate}$$

As shown in table 9, GradRate has a positive and statistically significant relationship with the percentage of alumni who donate. The coefficient is 0.3541 with a p-value of 0.000, meaning that

as the graduation rate increases, alumni donation percentages also tend to rise. This could suggest that students who successfully complete their education are more satisfied and therefore more likely to give back. In table 10, the R-squared value is 0.241, showing that about 24.1% of the variation in alumni donations can be explained by graduation rate alone, which is pretty good for single-variable models.

percalumni -> Expend + Top10perc

Variable	Coefficient	Standard Error	p-value
Intercept	11.8275	0.840	0.000
Expend	0.0005	9.96e-05	0.000
Top10perc	0.2238	0.029	0.000

Table 11: Regression Results for Alumni Donation Percentage using Expenditures and Top 10% Students

Statistic	Value
R-squared	0.232
Adjusted R-squared	0.230
F-statistic	116.7
Prob (F-statistic)	5.18e-45
AIC	5917.0
BIC	5931.0
Durbin-Watson	1.819

Table 12: Model Summary Statistics for OLS Regression with Expenditures and Top 10% Students

$$\text{percalumni} = 11.8275 + 0.0005 \times \text{Expend} + 0.2238 \times \text{Top10perc}$$

As shown in table 11, both Expend and Top10perc have positive and statistically significant associations with the percentage of alumni donations. The coefficient for Expend is 0.0005 with a p-value of 0.000, suggesting that higher per student expenditures are associated with increased alumni giving. Similarly, the coefficient for Top10perc is 0.2238, also with a p-value of 0.000, indicating that universities with a greater proportion of students from the top 10% of their high school class tend to receive more alumni donations.

In table 12, the R-squared value is 0.232, meaning that approximately 23.2% of the variation in alumni donation rates can be explained by these two variables combined. This represents an improvement compared to the simpler models. When Expend was the only predictor, the R-squared was 0.174, and when Top10perc was, the R-squared was 0.207.

percalumni -> Expend + Top10perc + Private

Variable	Coefficient	Standard Error	p-value
Intercept	7.0792	0.896	0.000
Private[T.Yes]	9.2050	0.845	0.000
Expend	0.0003	9.48e-05	0.003
Top10perc	0.2265	0.027	0.000

Table 13: Regression Results for Alumni Donation Percentage using Private Status, Expenditures, and Top 10% Students

Statistic	Value
R-squared	0.334
Adjusted R-squared	0.331
F-statistic	129.2
Prob (F-statistic)	7.96e-68
AIC	5808.0
BIC	5826.0
Durbin-Watson	1.994

Table 14: Model Summary Statistics for OLS Regression with Private Status, Expenditures, and Top 10% Students

$$\text{percalumni} = 7.0792 + 9.2050 \times \text{Private[T.Yes]} + 0.0003 \times \text{Expend} + 0.2265 \times \text{Top10perc}$$

As shown in table 13, all three variables, Private, Expend, and Top10perc, have positive and statistically significant associations with alumni donation percentage. The coefficient for Private[T.Yes] is 9.2050 with a p-value of 0.000, indicating that private institutions tend to have alumni donation rates approximately 9.2 percentage points higher than public ones. Expend has a coefficient of 0.0003 and a p-value of 0.003, showing that higher spending per student is still positively associated with alumni giving, though the effect size is smaller compared to simpler models. The coefficient for Top10perc remains strong at 0.2265 with a p-value of 0.000, reaffirming the importance of student academic quality.

In table 14, the R-squared value is 0.334, meaning that about 33.4% of the variation in alumni donation rates can be explained by university type, expenditures, and student quality combined. This represents a notable improvement in explanatory power over previous models. Private alone explained only 7.1% of the variation, while Expend and Top10perc combined explained 23.2% . The inclusion of all three predictors offers a more complete understanding of what drives alumni giving.

percalumni -> Expend + Top10perc + Private + Terminal

Variable	Coefficient	Standard Error	p-value
Intercept	-1.6639	2.276	0.465
Private[T.Yes]	10.2636	0.874	0.000
Expend	0.0002	9.65e-05	0.053
Top10perc	0.1898	0.029	0.000
Terminal	0.1243	0.030	0.000

Table 15: Regression Results for Alumni Donation Percentage using Private Status, Expenditures, Top 10% Students, and Terminal Degree Proportion

Statistic	Value
R-squared	0.349
Adjusted R-squared	0.345
F-statistic	103.3
Prob (F-statistic)	1.89e-70
AIC	5792.0
BIC	5816.0
Durbin-Watson	1.938

Table 16: Model Summary Statistics for OLS Regression with Private Status, Expenditures, Top 10% Students, and Terminal Degree Proportion

$$\text{percalumni} = -1.6639 + 10.2636 \times \text{Private[T.Yes]} + 0.0002 \times \text{Expend} + 0.1898 \times \text{Top10perc} + 0.1243 \times \text{Terminal}$$

As shown in table 15, the model includes four predictors of alumni donation percentage: Private, Expend, Top10perc, and Terminal. All variables except the intercept are positively associated with alumni giving, and most are statistically significant. The coefficient for Private[T.Yes] is 10.2636 with a p-value of 0.000, confirming that private institutions tend to have significantly higher alumni donation rates. Top10perc remains a strong predictor with a coefficient of 0.1898 and a p-value of 0.000, indicating that academically stronger students are linked to greater alumni support. Terminal, which measures the percentage of faculty with terminal degrees, also shows a positive and statistically significant effect, suggesting that faculty qualifications contribute meaningfully to alumni engagement. The coefficient for Expend is 0.0002 with a p-value of 0.053, which indicates an insignificant relationship with alumni giving.

Table 16 reports an R-squared value of 0.349, meaning that 34.9% of the variation in alumni donation rates is explained by this four-variable model. This is the highest R-squared among all models examined, showing improved explanatory power over previous combinations. For example, Terminal alone accounted for just 7.1% of the variation, and the model with Private, Expend, and Top10perc explained 33.4%. The addition of Terminal provides better model.

percalumni -> Expend + Top10perc + Private + Terminal + GradRate

Variable	Coefficient	Standard Error	p-value
Intercept	-7.4619	2.365	0.002
Private[T.Yes]	8.3449	0.893	0.000
Expend	0.0002	9.38e-05	0.050
Top10perc	0.1276	0.029	0.000
Terminal	0.0962	0.029	0.001
GradRate	0.1708	0.025	0.000

Table 17: Regression Results for Alumni Donation Percentage with Private Status, Expenditures, Top 10% Students, Terminal Degree Proportion, and Graduation Rate

Statistic	Value
R-squared	0.386
Adjusted R-squared	0.382
F-statistic	97.09
Prob (F-statistic)	2.40e-79
AIC	5748.0
BIC	5776.0
Durbin-Watson	1.995

Table 18: Model Summary Statistics for Full OLS Regression Model

As shown in table 17, all five predictors are positively associated with alumni donation percentage, and most of them are statistically significant. Private[T.Yes] has a coefficient of 8.3449, indicating that private institutions tend to have alumni donation rates approximately 8.3 percentage points higher than public institutions. GradRate is also a strong predictor with a coefficient of 0.1708 and a p-value of 0.000, reinforcing the idea that institutions with higher graduation rates foster greater alumni giving. Top10perc and Terminal remain significant and positive, reflecting the continued importance of student quality and faculty credentials. Expend is marginally significant, suggesting a small but positive impact of per student spending.

According to table 18, the R-squared value for this model is 0.386, indicating that approximately 38.6% of the variation in alumni donation rates can be explained by these five institutional and academic factors. This is the highest explanatory power achieved among all the models tested. For comparison, GradRate alone explained 24.1% of the variation, and the previous four-variable model explained 34.9%. Adding GradRate enhances the model's performance, suggesting the importance of graduation rate.

percalumni -> Top10perc + Private + Terminal + GradRate

Variable	Coefficient	Standard Error	p-value
Intercept	-7.8647	2.361	0.001
Private[T.Yes]	8.7698	0.868	0.000
Top10perc	0.1563	0.025	0.000
Terminal	0.1095	0.028	0.000
GradRate	0.1710	0.025	0.000

Table 19: Regression Results for Alumni Donation Percentage with Private Status, Top 10% Students, Terminal Degree Proportion, and Graduation Rate

Statistic	Value
R-squared	0.383
Adjusted R-squared	0.380
F-statistic	120.0
Prob (F-statistic)	1.39e-79
AIC	5750.0
BIC	5773.0
Durbin-Watson	1.982

Table 20: Model Summary Statistics (Excluding Expenditure Variable)

Expend was removed from the model because it was not statistically significant enough to justify its inclusion, showing a p-value of approximately 0.050 in the previous model.

As shown in table 19, the regression model that includes Private, Top10perc, Terminal, and GradRate all have positive and statistically significant relationships with alumni donation. The coefficient for Private[T.Yes] is 8.7698 with a p-value of 0.000, suggesting that private institutions have alumni donation rates approximately 8.8 percentage points higher than public institutions. Top10perc remains a significant predictor, indicating that schools with a higher percentage of top-performing students are associated with higher alumni giving. Terminal also shows a positive impact, suggesting that more highly qualified faculty contribute to higher alumni donations. GradRate continues to be a strong predictor, reinforcing that schools with higher graduation rates tend to see more alumni giving.

This model, in table 20, has an R-squared value of 0.383, meaning that approximately 38.3% of the variation in alumni donations can be explained by these four variables. This is very close to the performance of the five-variable model that included Expend, which had an R-squared of 0.386.

The model's p-value confirms that the predictors provide a highly significant explanation of the variation in alumni donation rates. Despite the removal of Expend, the adjusted R-squared value (0.380) indicates that this simplified model still accounts for most of the variance.

Conclusion

Overall, the model predicting percalumni using the five predictors(Expend, Top10perc, Private, Terminal, and GradRate) had an R-squared of 0.386, while the simpler model without Expend (Top10perc, Private, Terminal, GradRate) had an R-squared of 0.383. These two models performed almost identically, so I think both are solid options. Since the difference is very small, the simpler model could definitely be preferred if we want a more simpler approach. In addition, there are still other things worth exploring, such as checking for multicollinearity using VIF and trying out different combinations of predictors to see if we can improve the model further.