

Statistical Research in Institutional Tutoring Data

Dawson I. Damuth

State University of
New York at Oswego

May 2024

Outline

1 Introduction

- Goals and Objectives
- Project Hardships

2 Methodology

- Applied Techniques
- Why Logistic Regression?

3 Results

- Model Building & Diagnostics

4 Conclusion

5 References

Introduction

- Analyzing tutoring records to see if there's a notable effect on course grades due to embedded tutor presence.
- We will perform a full-scale comprehensive analysis of our data using select methods.
- "Is It the Intervention or the Students?".¹

¹Roddy Theobald and Scott Freeman

www.lifescied.org/doi/10.1187/cbe-13-07-0136.

Methodology

- Logistic Regression and assumptions of model building.
- Alternative methods of analysis.
- Final model assessments.

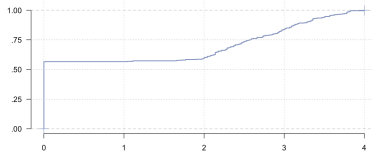
Results

When analyzing our data we can calculate the following frequencies:

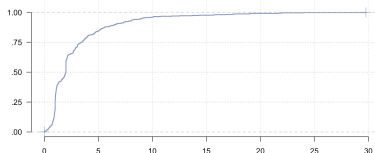
- Course Passed (Response): 64%
- Embedded Tutor: 22.3%
- Previously Attempted: 50.6%

We can also produce descriptive statistics for our continuous predictor variables:

- Prior GPA (4.0 Scale): $\mu = 1.2$, $\sigma = 1.43$



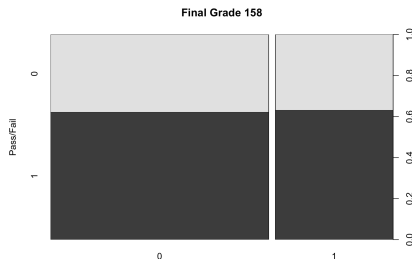
- Duration (Hours): $\mu = 2.98$, $\sigma = 3.59$



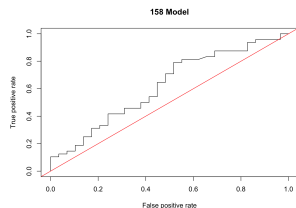
Model Building

Our model for MAT158 provides the following changes to the odds of a student's successful completion:

- Embedded tutor presence → 22.4% decrease
- Increase in prior GPA (+1.0) → 22.9% increase
- Increase in duration (+1hr) → 12.5% increase
- Additional previous attempts → 21.9% increase



The confusion matrix for our model is determined by Receiver Operating Characteristic (ROC) curve analysis:



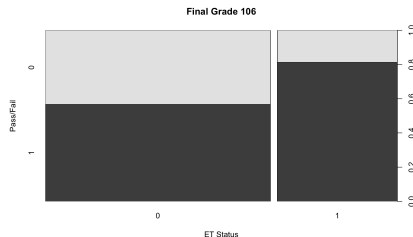
Prediction, $\pi_0 = 0.56$

Actual	$\hat{y} = 0$	$\hat{y} = 1$	
y=0	TN = 14	FP = 15	29
y=1	FN = 10	TP = 38	48
Total	24	53	77

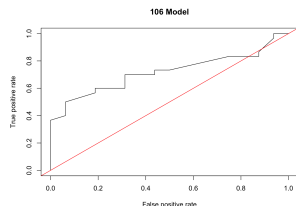
Accuracy: 65.22% Sensitivity: 50.00% Specificity: 93.75%

Our model for MAT106 provides the following changes to the student's odds:

- Embedded tutor presence \rightarrow 226.6% increase
- Increase in prior GPA (+1.0) \rightarrow 53.5% increase
- Increase in duration (+1hr) \rightarrow 16.7% increase
- Additional previous attempts \rightarrow 27.7% increase



For our MAT106 model:



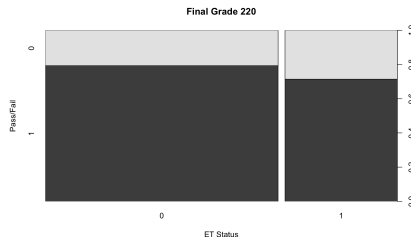
Prediction, $\pi_0 = 0.76$

Actual	$\hat{y} = 0$	$\hat{y} = 1$	
y=0	TN = 15	FP = 1	16
y=1	FN = 15	TP = 15	30
Total	30	16	46

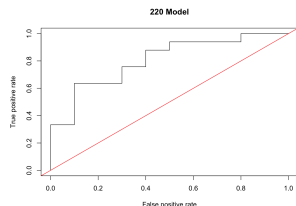
Accuracy: 67.53% Sensitivity: 79.17% Specificity: 48.28%

Our model for MAT220 provides the following changes to the student's odds:

- Embedded tutor presence \rightarrow 62.2% decrease
- Increase in prior GPA (+1.0) \rightarrow 28.7% decrease
- Increase in duration (+1hr) \rightarrow 30.5% increase
- Additional previous attempts \rightarrow 1022.2% increase



For our MAT220 model:



Prediction, $\pi_0 = 0.85$

Actual	$\hat{y} = 0$	$\hat{y} = 1$	
y=0	TN = 9	FP = 1	10
y=1	FN = 12	TP = 21	33
Total	19	22	43

Accuracy: 69.77% Sensitivity: 63.64% Specificity: 90.00%

Conclusions

- Model implications and difficulties.
- Future analysis.
- Learning outcomes.

Thank you!

Any Questions?

References

Roddy Theobald, Scott Freeman
www.lifescied.org/doi/10.1187/cbe-13-07-0136.