Problem 54-2

awesomedigger1021

October 2020

Introduction

Data:

Martha was accepted. She was the 95th percentile of her class, got a 33 on the ACT, and had an internship at a well-known company the summer before she applied to college.

Jeremy was rejected. He was in the 95th percentile of his class and got a 34 on the ACT.

Alphie was accepted. He was in the 92nd percentile of his class, got a 35 on the ACT, and had agreed to play on the college's basketball team if accepted.

Dennis was rejected. He was in the 85th percentile of his class, got a 30 on the ACT, and had committed to run on the college's track team if accepted.

Jennifer was accepted. She was in the 80th percentile of her class, got a 36 on the ACT, and had a side business in 3D printing that was making 15,000 dollars per year.

Martin was rejected. He was in the 85th percentile of his class, got a 29 on the ACT, and had was a finalist in an international science fair.

Mary was accepted. She was in the 95th percentile of her class, got a 36 on the ACT, and was a national finalist in the math olympiad.

Dean was rejected. He was in the 87th percentile of his class, got a 31 on the ACT, and was a national finalist in the chemistry olympiad.

Adam was accepted. He was in the 99th percentile of his class and got a 36 on the ACT.

Jeremy was rejected. He was in the 95th percentile of his class and got a 32 on the ACT.

1 Create a quantitative dataset to represent this information, and include it in your writeup. Name your features appropriately.

Data:

[Name,Percentile,Act score,Extracurricular,Accepted]

[Marsha, 95, 33, Internship, True]

[Jeremy, 95, 35, None, False]
[Alphie, 92, 35, Baseball, True]
[Dennis, 85, 30, Track, False]
[Jennifer,80, 36, Business, True]
[Martin, 85, 29, Science fair, False]
[Mary, 95, 36, Math Olympiad,True]
[Dean, 87, 31, Chem Olympiad,False]
[Adam, 99, 36, None, True]
[Jeremy, 95, 32, None, False]

2 Decide what type of model you will use to model the probability of acceptance as a function of the features in your dataset. State and justify the form of the model in your writeup

I will be using a logistic model because of our need for a number between one and zero and the necessity for interactions with our variables. We want a number between one and zero to show the percent chance of us getting into college. My terms will be Percentile, Act score, and Extracurricular. We need interaction to make up for the fact that you can come from a small school and have be in a high percentile, and being in an extracurricular can help make up for not great testing scores or low percentile, but not both.

3 Fit the model to the data.

For each feature, answer the following questions:

According to your model, as that variable increases, does the estimated probability of acceptance increase or decrease? Does that result make sense? If so, why? (If not, then something is wrong with your model, and you need to figure out what's going wrong.)

Percentile Term:
Your acceptance probability increases as your Percentile increases. This makes sense because having a high percentile means you're a good student in Theory.

Act score Term:
Your acceptance probability increases as your Act score increases. This makes sense because it's a good wait to tell your intelligence(disregarding my own personal beliefs on testing and its relation to intelligence in our flawed education system) regardless of other biased situational quirks.

Extracurricular Term:
Your acceptance probability increases if you did an Extracurricular. This makes

sense because it's a nice added bonus if you also dont have an awful Percentile and Act score.

Percentile_Act Score Term:
Your acceptance probability decreases as your Percentile_Act Score increases. This makes sense because it acts as a precaution against a biased Percentile score and because collages really want well rounded students that do more than the base curricular. This adds stress for Extracurricular activities.

Percentile_Extracurricular Term:
Your acceptance probability decreases as your Percentile_Extracurricular Term increases. This makes sense because it's a suggestion that your Percentile may be less accurate due to how difficult it is to have high grades and do Extracurricular activities. This is just to take off some of the bias caused from coming from a small school.

Act score_Extracurricular Term:
Your acceptance probability increases as your Act score_Extracurricular increases. This makes sense because if you do an Extracurricular, you're more desirable and so your Act score will be weighted heavier.

4 Estimate the probability of being accepted for each of the data points that you used to fit the model. How well does this match up with reality?

Martha: 0.9966895600782484
Jeremy: 0.0009999946440241599
Alphie: 0.9979681707761708
Dennis: 0.000986149161169717
Jennifer: 0.9983507182902432
Martin: 0.00010058029388191834
Mary: 0.9994654728267108
Dean: 0.056700032889175724
Adam: 0.9989999940672725
Jeremy: 0.0009999949836823289

This matches well with reality, if they got above %50 then they were accepted, and if not, they weren't.

5 Estimate your probability of being accepted if you are in the 95th percentile of your class and got a 34 on the ACT. Justify why your model's prediction is reasonable.

David: 0.0009999946440241599 would be the probability. This makes sense because I only have a mild Act score and Percentile ranking, and the only way we've seen someone with stats like that get in is with an extracurricular, kinda like Alphie.

6 Now suppose that you have an opportunity to do an internship at a well-known company the summer before you apply to college. If you do it, what will your estimated probability of acceptance become? Based on this information, how much does the internship matter in terms of getting into the college you want?

The new probability would be David: 0.9981962881896385, this confirms what I said in part e, the internship would be the deciding factor for me getting into the collage.