

Modeling Workspace

William Morgan

11 April, 2018

Problem Outline

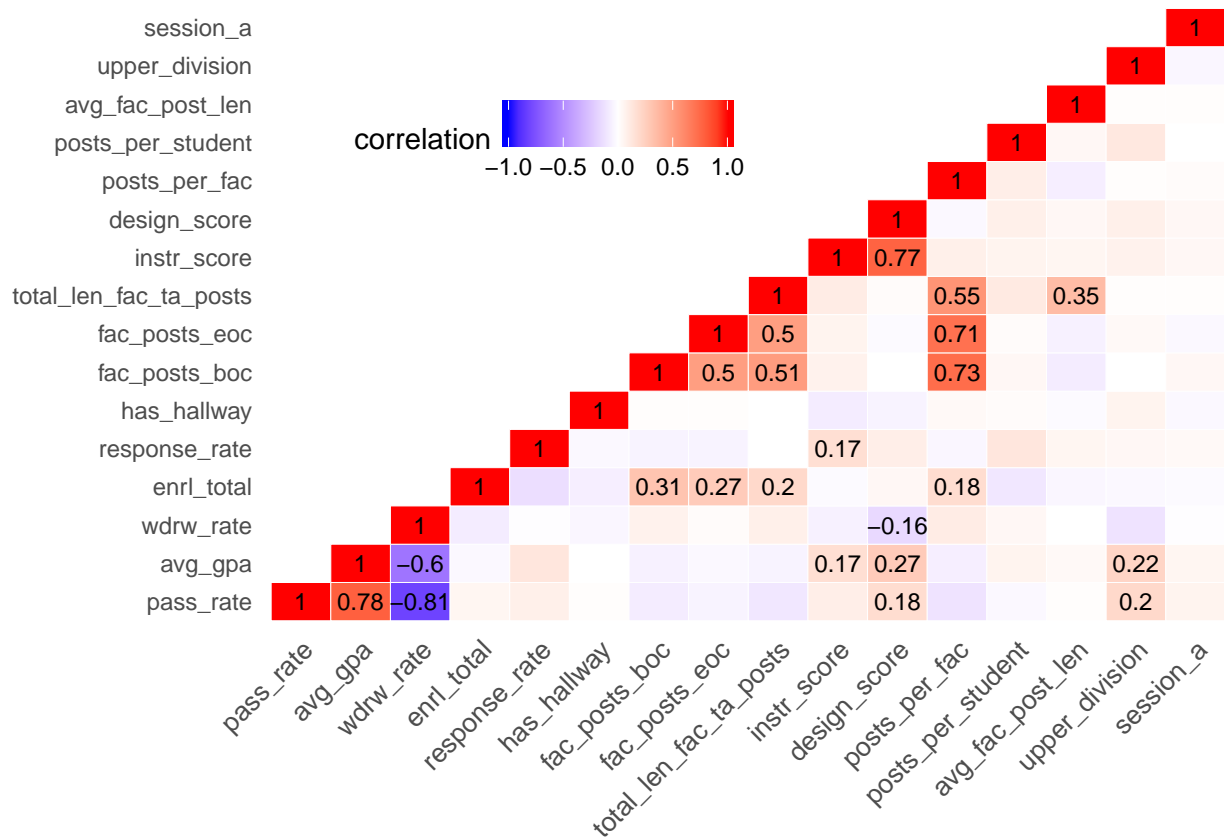
What do we want to answer?

- *Q1*: Does faculty activity on discussion forums affect student engagement on forums?
 - *Q2*: How do student outcomes change with the level of engagement in online forums? If there is a change, it is independent of who is doing the posting?
 - *Q3*: Is the importance of faculty engagement constant throughout the course or are there specific points in time where students benefit more from faculty engagement?
-

Preliminary Feature Search

We have some expectations on which variables might carry the most signal, but it is worth doing some investigation in case there is something we overlooked. Our first objective will be creating a correlation heat map of all the variables that we anticipate using, including the responses we wish to test.

This section will be relatively brief, as we prefer to avoid a deep dive into the potential feature space. Instead, we will adopt methods that help us in selecting appropriate features.



Research Question 1

Does faculty activity on discussion forums affect student posting behavior?

Based on the available data, we can use several metrics for student behavior on discussion forums:

- number of posts per student (over the entire course)
- average length of posts made by students
- consistency of student posts (from week to week)
- variance of the number of posts in week X

Notes:

We need to be careful about what we choose to include in these models because some things might not be interpretable in a meaningful way. If we were to use `pass_rate` as a predictor and find that it has a positive effect on say the number of posts per student, what would our interpretation be? That increasing the pass rate of a course creates more student discussion? That's silly and feels like some sort of logical fallacy.

- generally the problem is: A is caused by B even though B occurs after A

With that said, we can still use things that are fixed or occur during the course. Class sizes, existence of hallway forums, faculty activity, and others can all be a part of this question.

Another caveat with this question is that our sample won't depend on missing values for the course survey data. This should add in at least 2100 new observations, given that those also do not have missing data for

any of the other variables we are using.

Future Options

- can we add data about instructors making announcements?
 - what peoplesoft data can be used?
-

Working Notes

- research questions should be reordered; train of thought should be:
 - does faculty activity on forums affect grade outcomes?
 - what characteristics of faculty activity have greatest effect? (early course posts, late course posts, consistency of posting)
 - does faculty activity on forums affect student activity on forums?
 - what student activity do we want to measure?
-

To-Do:

- sample selection needs to be dynamic; this is probably best resolved after developing Q2, Q3 more so you know exactly what sample is needed in each
- decide on new variables that need to be created and stick to it; no need to go crazy
- decide if you're going to extend the prelim feature analysis