Midterm 1, Part 2

S&DS 361

2023-02-21

For this part of the midterm, you can use R and R Studio, class notes, Google, ChatGPT, and the internet in general (e.g Stack Overflow). You may not use Ed Discussion or any other messaging/discussion forum. You may not communicate directly with any other human being. Of these resources, I'm guessing R and R Studio will be useful and the others will not be that useful.

Academic Salaries

This question involves the `academic.salaries.title.gender.rds` data set posted on GitHub here under `exams/midterm1/data`.

Build a linear regression model using the `academic.salaries.title.gender.rds` data to find the expected `salary` for a professor at a public university based on other information known about that professor. Only consider the columns `salary` (in dollars), `group.title` (Assistant Professor, Associate Professor, etc), `male` (1=male, 0=female), `score` (from US News and World Report rankings), `region`, and `state`, but feel free to transform any of those columns if you think it would help. Try a few models. For each model you try, handwrite the outcome, predictors, and Adjusted R^2 in a table like this:

```
   Outcome       Predictors AdjustedR2
 myoutcome          mypred1      0.500
 myoutcome mypred1, mypred2      0.550
```

Handwrite that table here:

Which model would you choose as the best model? Give a short explanation (1-2ish sentences).