

Assignment 2

Lacey Hartigan

due 1/26/21

For this assignment, you'll be working with a county-level dataset (`pd.Rdata`) to predict home ownership rates using conditional means. You'll need to select the county-level characteristics that you think might be related to home ownership rates.

Submit your assignment as `assignment2_<yourlastname>.Rmd`, where `<yourlastname>` is your last name. (By the way, any time you see this: `<sometext>`, that indicates that you need to substitute something in, so if I were to submit the above assignment, it would be as: `assignment2_Hartigan.Rmd`). Also, be sure to submit *ONE* knit file (in the format of your choosing). I have inserted R code chunks for you below; however, in future assignments I won't do this step for you.

I expect that the `.Rmd` file you submit will run cleanly, and that there shouldn't be any errors. Use LOTS of comments to tell me what you are doing. Also, please update the "author" in the header to be you. Finally, identify the students with whom you worked (there is a space for this at the end of the assignment).

1. Calculate the mean of the outcome `homeown_rate`.
2. Use your mean as a prediction: Create a new variable that consists of the mean of the outcome.
3. Calculate a summary measure (RMSE) of the errors for each observation—the difference between your prediction and the outcome.
4. Calculate the mean of the outcome at levels of a predictor variable *of your choosing* (make sure it makes sense as a predictor for home ownership).
5. Use these conditional means as a prediction: for every county, use the conditional mean to provide a "best guess" as to that county's level of the outcome.
6. Calculate a summary measure of the error (RMSE) in your predictions.
7. Did your prediction improve? How do you know whether or not it improved?

Study buddies for this assignment (maximum of three partners):