Problem set 6

due Monday, December 9, 2024 at noon

> 💡 Run your code
>
> Make sure your code can run before submission! Runtime > Run all

Instructions Upload your `.ipynb` notebook to gradescope by 11:59am (noon) on the due date. Please include your name, Problem set number, and any collaborators you worked with at the top of your notebook. Please also number your problems and include comments in your code to indicate what part of a problem you are working on.

Problem 1

This problem set deals with the stroop data we worked with in Tuesday's class. Here is a demo of the experiment. Explore these data with (at least) glimps and a ggplot appropriate for the data. Include accuracy as the response variable (output) and condition as the explanatory variable. You may explore the data in any other ways you see fit.

Problem 2

Specify a model (with an equation) that predicts accuracy from condition. Then, fit the model appropriately (with base R, infer, or parsnip). Make sure to return the parameter estimates.

Problem 3

Estimate the accuracy of the specified model on the population using k-fold cross validation. Use the `collect_metrics()` function to return the R^2 value.

Problem 4

Using bootstrapping with `infer` (at least 1000 reps) to quantify your model's reliability. Get a 95% confidence interval the percentile way. Then visulize your sampling distribution with `visualize()`, shading the confidence interval in green.

Problem 5

Suppose you want to add RT as a predictor in your model. Specify (with an equation) and fit the model with `lm()` (or `glm()` if appropriate). Then use the `check_collinearity()` function in easystats to check for multicolinearilty in the model's preditors. In a text cell, write whether they are correlated (based on the output of `check_collinearity()`) and what that means for your model.

Problem 6

The stroop data actually includes *repeated measures*, wherein each subject completes multiple trials. Explore the data again and choose a visualization that can incorporate `subj`. Specify (with an equation) and fit a mixed effect logistic regression model predicting accuracy by condition and RT (fixed effects) with a random intercept for `subj`. Make sure to return the parameter estimates.

Challenge (optional)

Suppose this was your own experiment. Re-explore the data and specify and fit the model you'd choose to model accuracy. Which fixed effects and why? Random interecpts for subjects, random slopes, or both? Explain why you made these decisions (based on your exploratory data analysis, model accuracy measurements, etc.).

No points awarded or removed for this question! Just for fun!