Problem set 7

due Monday, Nov 3, 2025 at 11:59am (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 uses the Stroop dataset we worked with in class. The Stroop task measures how quickly people can name the color of the ink that a word is printed in while ignoring the word itself. A demonstration of the experiment is available here.

Begin by filtering the data to include **only trials where participants responded accurately**. Then explore the data using `glimpse()` and a visualization created with `ggplot`. Include **reaction time (RT)** as the response variable and **condition** as the explanatory variable. You may explore the data in other ways as well — for example, by comparing the distribution of RTs across conditions or by looking at potential outliers.

Problem 2

Next, specify a model that predicts **reaction time** from **condition** for accurate trials. Write your model as an equation and describe what each part represents. Fit this model appropriately using either `lm()`, `infer`, or `parsnip`. Return the parameter estimates and interpret them in your own words. What does your model suggest about how the Stroop condition affects participants' speed of responding?

Problem 3

Evaluate how well your model generalizes to the population using **k-fold cross-validation**. Use the `collect_metrics()` function to return the R^2 value. Report this value and describe what it tells you about how much of the variation in reaction time your model explains. What would a high or low R^2 mean in this context?

Problem 4

Use **bootstrapping** with `infer` (at least 1000 replications) to estimate the reliability of your model. Construct a 95% confidence interval for one of your model parameters using the percentile method. Visualize your bootstrap sampling distribution with `visualize()`, shading the confidence interval in green. In a brief text response, describe what this confidence interval tells you about the stability of your model's estimates.