Problem Set 8

Due Monday, Nov 17 at 11:59am (noon!)

Instructions

Upload your `.ipynb` notebook to **Gradescope** by **12:00 PM (noon)** on the due date. At the top of your notebook, include your **name**, **problem set number**, and any **collaborators**. Please number your problems and include comments in your code to indicate what part of a problem you are working on.

Problem 1

In this problem set, you will return to the Stroop dataset — the same one you analyzed in the previous problem set — but now you'll model **accuracy** as a categorical outcome instead of **reaction time** as a continuous one. This means you'll be fitting a **classification model** instead of a **linear model**.

The Stroop task measures how quickly and accurately 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 exploring the data with `glimpse()` and a visualization using `ggplot`. Include **accuracy** (correct vs. incorrect) as the response variable and **condition** as the explanatory variable. You may explore in any other ways you find useful, such as plotting the proportion of correct responses across conditions.

Problem 2

Specify a **classification model** that predicts whether a response is correct based on **condition**. Write your model as an equation and explain what each part represents. Fit this model using a logistic regression approach — for example, with `glm()` in base R, with `parsnip`, or within an `infer` workflow. Return the parameter estimates and interpret them in words. What do the model's coefficients suggest about the relationship between Stroop condition and the probability of an accurate response?

Problem 3

Estimate your model's performance on the population using **k-fold cross-validation**. Use the `collect_metrics()` function to return both **classification accuracy** and **ROC-AUC** values. Report these values and describe what they tell you about how well your model performs. Which metric is more informative for this kind of task, and why?

Problem 4

Use **bootstrapping** with `infer` (at least 1000 replications) to quantify the reliability of your classification model. Construct a 95% confidence interval for one of your model parameters (for example, the difference in log-odds between the congruent and incongruent conditions) using the percentile method. Visualize your bootstrap sampling distribution with `visualize()`, shading the confidence interval in green. In a short text response, explain what the width of this interval tells you about the stability of your model's estimates.

Problem 5

Suppose you want to test whether participants are more likely to respond correctly when they respond **more quickly**. Add **reaction time (RT)** as a predictor in your model and fit a **multiple logistic regression**. Then use the `check_collinearity()` function from the **easystats** package to test whether `condition` and `RT` are correlated. In a short text response, describe what you find. If the predictors are correlated, what does that mean for how you should interpret their effects in the model?