Background & Introduction

- Aphasia is a communication disorder that results from damages to parts of the brain that controls language and speech.
- It usually happens after a sudden stroke or brain injury.
- Depending on which part of the brain that’s injured, this results in different types of Aphasia. The disorder can range from mild to severe.
- The WAB Aphasia Quotient Score determines which Aphasia group each patient belongs to.

Exploratory Data Analysis

Data / Data Processing

- There are 5 datasets that were collected through FLUCALC (an automated language analysis tool). Each dataset consists between 202-559 data points and has been recorded through various aphasia tests.
- Each of the datasets can be categorized into three of the following tasks: narrative, expositional, and procedural.
  - All the plots are based on the Cinderella (narrative) task. The Cinderella task is performed by asking the patients to retell the story of Cinderella and record their response.
  - The following predictor variables are listed: log % of phrase repetitions and log % word revisions.
  - There were a handful of anomalies in the dataset; most of them were related to the specific aphasia groups (missing data/incorrectly labeled data).

Boxplots for Cinderella Task Dataset

- The boxplots above show the distribution across all aphasia groups for each predictor variables.
  - The predictor variables were all log transformed in order to compare the medians of each group more properly.
  - Figure 1 has the predictor “log % of phrase repetition” on the y-axis. It shows that the medians across all aphasia groups are different. The control group has the lowest median, while Conduction and Wernicke (2 of the most severe types of Aphasia based on the WAB AQ score) have the highest medians.

Analysis & Results

Analyzing Patterns Across Aphasia Types: ANOVA Tests

- Since we obtain a p-value of < 0.05 after running the ANOVA tests, we conclude that at least one Aphasia group has a significantly different mean than that of the other Aphasia groups for all the important variables (log transformed and added a constant of 0.01) we selected after doing EDA.
- In Figure 3 and 4, we see the 95% confidence intervals of log % phrase repetitions and log % word revisions between pairwise Aphasia groups. The intervals that do not overlap the middle dotted line (mean = 0) suggest significant differences in the mean. For all significant variables, the pairs with significant differences include Broca - Anomic, Control - Anomic, and Control - Conduction.

Analyzing Patterns Across Aphasia Types: Principal Component Analysis

- PC1 captures 36.52% while PC2 captures 24.33% of the total variance in the data, based on figure 5.
- The predictors log number of utterances, log number of words, and log number of words per minute contribute to the mean most to the variance captured by PC1.
- We observe a cluster for the control group; it has positive PC1 values, whereas other groups have negative PC1 values. There are no identifiable patterns for PC2 by Aphasia groups (see figure 5).

Conclusions & Future Work

- We observe that some Aphasia groups have significantly different means for the predictors of interest.
- There might be a better way to more accurately classify the patients into each Aphasia group based on the specific variables we looked at, other than the WAB AQ score.
- In the future, we hope to contribute to the development of methods for quantifying fluency behaviors on a larger scale.
  - The data we used was collected from multiple different Aphasia labs, which could result in biases/inconsistencies in the dataset.
  - The results we have are only based on the Cinderella dataset; other tasks may yield different results.
- We would like to give special thanks to Dr. Joel Greenhouse, Dr. Davida Fromm, and Dr. Zach Branson for guiding us throughout the research process and providing us valuable feedback over the semester.