Top of Form

|  |  |  |
| --- | --- | --- |
| **Client:** | Leah Nodar | **File:** 24-050 |
| **Dept:** | Linguistics | **Faculty:**  **Student:** |
| **Date:** | 7/10/2024 | **Initial Meeting:**  **Follow-up:** |
| **Consultant and Attendees:** Sumeeth Guda, Leah Nodar, Dr. Felicia Roberts, Dr. Bruce Craig | | |
| **Statement of Problem:**  To determine whether groups of speakers from Mobile and Africatown, Alabama have significantly different speech patterns, and investigate how the dialects evolved over time using seven linguistic features. | | |
| **Goal of this Project:** PhD Dissertation | | |
| **Background:**  The client is a PhD student in the linguistics department who is exploring whether two groups of people from the early 20th century, who had very different English dialects, continue to have these differences in the 1970s and 1990s, or if their dialects have converged.  The two groups are the people of Africatown, AL and those of Mobile, AL. According to the client, this is linguistically interesting for two reasons. First, because Africatown has a unique dialect that has not been studied before, and second, because Africatown has documentation of a town founder's earliest stage of the dialect in the 1920s, so following this across later time periods may tell us something new about how dialects develop overall, with implications for how other languages change.  The client will not be conducting an experiment, but rather perform a social/historical analysis. The client’s data are constrained to the audio available to the client from the two time periods mentioned, with a small sample within the groups. The client has two time periods they are working under (1970s, 1990s), and they have three groups = {Group A: The descendants of the original Africatown founders, Group B: People who were born and raised in Africatown but are not direct descendants, Group C: people from Mobile, AL}.    The distribution of the groups is:  1970s :   * Group A: 2 people * Group B: 2 people * Group C: 7 people   1990s:   * Group A: 3 people * Group B: 4 people * Group C: 11 people   The way the client is currently gathering the data is by listening to each of the people’s audio and marking up / flagging seven linguistic variables (pronunciations or grammatical features). These seven variables were chosen based on previous research that analyzed the documentation of speech from the 1920s of a founder of Africatown.  The client indicated the features she was looking out for in her attached document “SevenFeatures”.  As an example, one feature is the pronunciation of "th" sounds as "d" (as in "dey, dese, dose" vs "they, these, those"). For each person, the client followed linguistic criteria to determine ~60 potential environments (places where this sound could occur) and then checked the audio and spectrogram for each of those environments to determine whether the person actually pronounced it with "th" or "d" (or neither).  The other features are similar: Based on linguistic criteria, determination of the environments in each person's speech where a feature could occur, followed by a check for whether it does occur. Generally, there are at least 50-100 potential environments per feature.  The client’s goal is to determine whether as a whole this collection of seven features indicates that the groups of speakers have significantly different speech patterns, with a broader goal of exploring how dialects evolve over time. | | |
| **Progress of project at this time:** Presently Collecting Data | | |
| **Relevant information presented at meeting:**  Client explained how they are counting how many times the instance happens for each individual, and then aggregating the collective groups. They collect the data through listening to audio recordings from the 1970’s and 1990’s for the 3 groups. Specifically, for each of the 7 features she is looking at the number of the environment occurs, and the number of times the speaker has spoken. The individuals in each of the client’s groups are representative of the group as a whole since as Dr. Roberts mentioned none of the individuals are from the same family.  As Dr. Craig recommended at the start, to compare the 70's and the 90's blocks, the client can look at and make comparisons with each feature. Specifically, through each proportion since the groups are different sizes. What she should do is make a comparison of the frequencies and tell if there is difference between the 70's and 90's. The client indicated in her application that she is concerned that her results might not be the most accurate because of her small sample size. Dr. Craig agreed with this as it might be tricky to determine significant differences because of the small sample size hence other methods such as logistic or binomial regression would have to be used to determine the differences. Another issue that was discussed was that since the client was working with such a small sample, it was imperative that her sample had to be representative of the groups they are from. | | |
| **Recommendations for Design and/or Analysis:**  Dr. Craig recommended that the client create a small dataset with a limit on the number of individuals from each group. For each individual have a row in an Excel spreadsheet and with the first column count the features’ frequencies. The next column will be the number of environments where they speak the features, the last column is the group labeling (A,B,C), and the final is what time period (70s, 90s). Using this approach the client can include groups, features, and interaction terms. The breakdowns of the groups can show how the groups are different empirically.  For statistical testing one approach which was recommended was to use was binomial regression. The only danger through using this model was that overdispersion can happen given that the sample size is small. However, since the binomial distribution probability of success defines the variance, the variance cannot change with the subjects. This is a problem, because there might be more than just binomial variability which occurs within the covariates of the groups such as gender and sex might have more variability. The variability of the covariates might not account for the additional variability under the binomial model. However, we cannot determine this to be the case until the model has been run.  The final recommendation for analysis was:   1. Client should pick one or two features to analyze individually first. Doing analysis on all 7 features at once will lead to a more complicated model, and in any case, she would have to analyze the features individually, and gradually work her way up to interaction effects. 2. The consultant should conduct binomial analysis first. to see each of the proportions / ratios for each individual and looking into the proportion differences. The binomial model will care about probabilities of differences of the features. If there exists any additional variability that could occur to over or under dispersion, then shift to a quasibinomial model to account for the dispersion. 3. The probability of success is rather small for the dataset. When this is the case, the Poisson distribution can be leveraged in this setting. Use a Poisson model to compare the distributions between the groups. It is important to note that this should be used only if there are any issues with the binomial model. | | |
| **Who will carry out these actions?**  Client:   * Finalize data collection, and add observations to an Excel spread sheet. * Share the data with the consultant along with any other relevant documents regarding the data collection methods and data contents. * Select 1 or 2 features to use for the analysis.   Consultant:   * Get access to the client’s data. * Create an R file containing all the model findings and run through each of the models for analysis. * Meet with the client to show the model results and explain how to conduct future analysis. | | |
| **Status:** Follow up meeting necessary | | |

Bottom of Form