**Assignment 3 - Diagnosing schizophrenia**

**Find .md files here:**

**Part 1&2 -** [**https://github.com/alexandermirz/alexandersgitnest/tree/master/Assignments/3**](https://github.com/alexandermirz/alexandersgitnest/tree/master/Assignments/3)

## **Part 1**

### **Exercise 1 - Data cleaning**

Upon writing a function, that could extract the participant and study data as well as doing the fundamental statistics on the data, we proceeded to merge the clinical data with duration- and pitch data. This was done only using the data from the Danish studies (1-4 - see script). Along with the merging, we encountered some omitting due to dublettes and missing values. This all resulted in the final data frame called “demo\_art\_pitch” which became the basis for the rest of the assignment.

### **Exercise 2 - Data description**

|  |  |  |
| --- | --- | --- |
| **Diagnosis-specific data** | Schizophrenic | Healthy |
| n() | 106 | 116 |
| Gender | M: 61 - F: 45 - 0.73 ratio | M: 66 - F: 50 - 0.75 ratio |
| Age (mean) | 26.49 years (SD 8.82) | 26.44 years (SD 8.96) |
| Education | 12.89 (SD 2.73) | 14.86 (SD 2.61) |
| SANS (neg. sympt) | 9.67 (SD 4.40) | 0.39 (SD 1.19) |
| SAPS (pos. sympt) | 10.33 (SD 4.91) | 0.08 (SD 0.43) |

Table 1 - sample description

The gender distributions are within-group not equal, but because the between group distribution thereof is almost perfect, we did not consider this an issue. Other than this, the groups seem pretty balanced in terms of age and education. Symptom severity is listed at the bottom as well as an indicator that there thankfully is a difference between the healthy and the diagnosed group.

When investigating the individual studies, the following results were obtained:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study/Diagnosis** | n() | Gender | Age (SD) | Edu (SD) | SANS (SD) | SAPS (SD) |
| 1/Healthy | 36 | M: 19 F: 17 | 22.66 (3.19) | 13.39 (2.19) | NA | NA |
| 1/Schizophrenic | 34 | M: 18 F: 16 | 22.84 (3.11) | 12.08 (2.31) | 10.27 (4.31) | 10.48 (4.03) |
| 2/H | 23 | M: 16 F: 7 | 23.65 (3.54) | 15.22 (2.56) | NA | NA |
| 2/S | 23 | M: 17 F: 6 | 23.34 (3.86) | 12.11 (2.50) | 9.94 (5.22) | 14.49 (4.32) |
| 3/H | 28 | M: 15 F: 13 | 37.69 (12.86) | 15.90 (2.65) | NA | NA |
| 3/S | 19 | M: 8 F: 11 | 41.05 (12.16) | 12.75 (2.69) | NA | NA |
| 4/H | 29 | M: 16 F: 13 | 24.36 (4.51) | 15.78 (2.17) | 1.29 (1.86) | 0.25 (0.75) |
| 4/S | 29 | M: 17 F: 12 | 24.76 (3.64) | 14.70 (2.61) | 8.61 (3.61) | 7.00 (3.91) |

Table 2 - sample description of each study

### **Exercise 3 - Data analysis and Result**

#### Analysis

We aimed to replicate the results of the meta-analysis’ 4 voice features which are characteristic of schizophrenia. Namely, a lower pitch variability, a lower proportion of spoken time, a slower speech rate, and a longer pause duration. Using lme4 package in Rstudio, we got estimates for all voice features having by diagnosis. There also several other factors in play - gender, age, and study. Ergo, we added gender and age as predictors. They turned out to be insignificant. When it comes to the study, things get more interesting. We assumed some minor differences in recording settings of every study. Therefore we scaled the relevant variables by means and standard deviations of each study to capture unnecessary noise. Also, we included a random intercept for participant’s id. The all 4 models looked as follows: voice feature ~ diagnosis + (1|ID) (where voice feature represents one variable at time).

#### Results

Below in table 3, our estimates of how the predictors are affected by diagnosis can be seen. Although the effects are somewhat lower, the **direction** of the effects seem to be the same. Speech rate is according to this the predictor that is most confidently predicted by diagnosis. This will thus be our best predictor from here on. Important to note is also, as above mentioned, that scaling has been performed on the data that we used (per. study). While the meta-analysis based the results on data from different languages, we based it only on Danish speaking participants which might also modulate the results.

|  |  |  |
| --- | --- | --- |
| Voice feature | Results of the meta-analysis (Hedge’s g) | **Our results** |
| pitch variability | -0.55, 95% CIs: -1.06, 0.09) | -0.226, 95% CIs: -0.367, -0.086 |
| proportion of spoken time | -1.26, 95% CIs: -2.26, 0.25) | -0.257, 95% CIs :-0.429, -0.086 |
| speech rate | -0.75, 95% CIs: -1.51, 0.04) | **-0.372**, 95% CIs: -0.536, -0.208 |
| pause duration | 1.89, 95% CIs: 0.72, 3.21 | 0.290, 95% CIs :0.143, 0.444 |

Table 3 - comparing results

## **Part 2 - Creating Automated Classifier**

#### Analysis

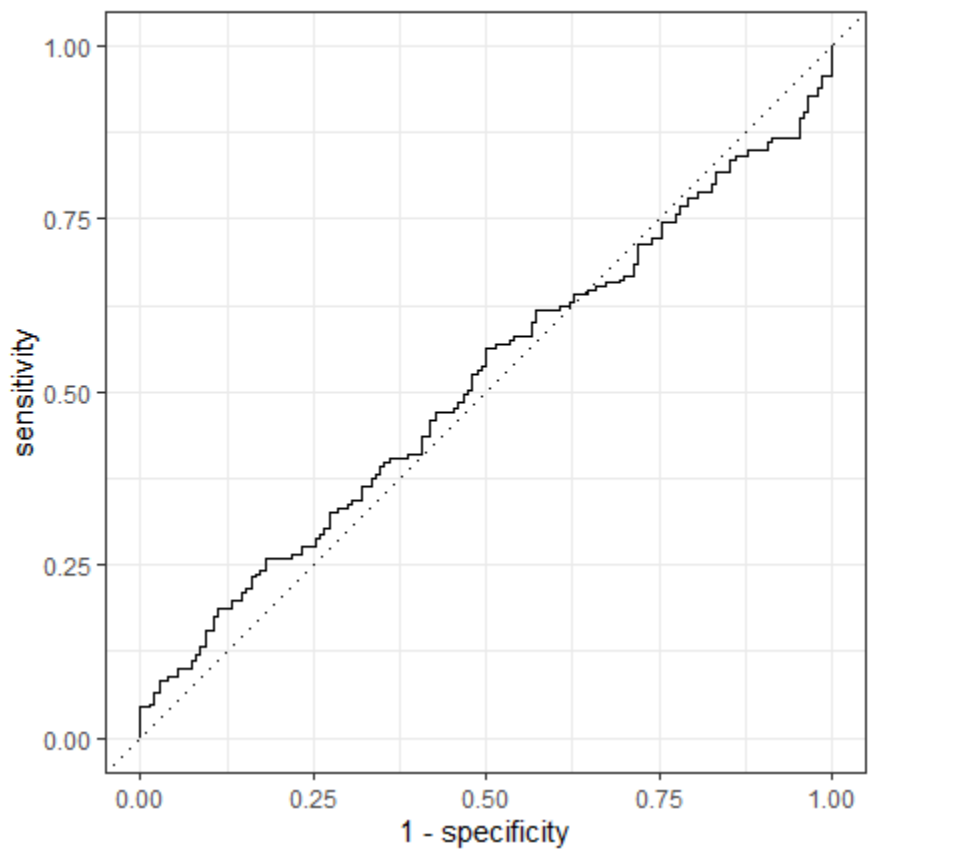
In part 2, we attempted to create an automated classifier of schizophrenia from the voice alone. To do so, we employed a logistic regression on our 1751 data points and 15 predictors (see appendix for list). The whole process was facilitated thanks to the authors of packages tidymodels (Kuhn & Wickham, 2019) and groupdata2 (Olsen L., 2019). However, in the tidymodels workflow is not possible to include random effects (yet). We cannot thereby account for any inter-study variability. The alternative approach is to scale per study (as in part 1) which would create additional incompatibilities with the tidymodels framework. As we ran out of options, we thus did not account for study-related variability. :-)

Should we include data from Japanese and Chinese studies as well, we would have run into similar problems but with the language. As the random-effect solution is ruled out, the separate analysis for every language would have to be conducted to account for the interlanguage variation. One might argue that the schizophrenic symptoms such as tendency for high-pitched voice should be detectable regardless of a language. And one is right, although the given effect size might slightly vary. Unfortunately, without specifying random effect for language, logistic regression would not consider this which would lead to unnecessarily skewed results.

#### Model with the best acoustic feature

Based on the analysis in the part 1, we identified the best acoustic feature a speech rate, given the highest effect size for diagnosing schizophrenia. Accordingly, we assume it to have the biggest predictive power. The accuracy of our logistic regression with the single feature is 0.477 (SD 0.096) thus it has no predictive power at all (see the roc curve below).

#### Model with the set of best acoustic features

Using logistic regression and all 15 relevant features, we yielded an accuracy of 0.514 (SD 0.225). Although it is slightly higher than for model with only speech rate feature, there is no improvement comparing this to merely chance. 

Nonetheless, using multimodal cross-validation, a more sophisticated method for training an algorithm, we obtained strikingly different results. We divided our training set into 10 fold with 100 repetitions of training sessions for every fold. The 17% increase in accuracy accuracy (0.68, SD 0.03) was reached when using a merely different training method. The accuracy is above the odds but still far from great nor potentially applicable for diagnosing. Cohen’s kappa, a more conservative measure which takes into account the aspect of randomness, shows very poor score of 0.36 (0.07). However, such measure is primarily used when the control and diagnosed subjects are imbalanced. Which is not our case and it might be too strict. The roc au score, a measure which considers a specificity/sensitivity trade-of, is 0.76 (SD 0.04), a too optimistic result.

*Cross validation model plotted*

Strengths and limitations of using logistic regression

The logistic regression is a simple but still highly efficient algorithm primarily used in machine learnings. It As long as the number of features is kept to minimum (not in 1000s), the results are stable and resistant to overfitting. Therefore, the feature engineering is for the logistic regression crucial. On the other hand, there are a lot of even more powerful algorithms which, in addition, can be used for non-linear relationships, contrary to the logistic regression. We tried different ML techniques including: random forest, support vector machine and boost tree. The results of these are listed in table 4.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of algorithm** | Accuracy (SD) | Cohen’s Kappa (SD) | ROC AU (SD) |
| logistic regression | 0.68 (0.03) | 0.36 (0.07) | 0.76 (0.04) |
| support vector | 0.71 (0.03) | 0.42 (0.07) | 0.78 (0.04) |
| random forest | 0.76 (0.03) | 0.52 (0.07) | 0.85 (0.03) |
| boost tree | 0.74 (0.06) | 0.47 (0.13) | 0.82 (0.06) |

Upon analyzing the data from healthy and schizophrenic participants we hereby conclude, that our models perform better than pure chance, but the accuracy seems not to be satisfying enough to apply in a clinical setting. Subsequently, we were not capable of replicating the effect size obtained by the meta-analysis - only the directionality. Possible pitfalls in our analysis could have been the amount of data (we sorted out all non-Danish),

### References

Alexandros Karatzoglou, Alex Smola, Kurt Hornik, Achim Zeileis (2004). kernlab - An S4 Package for Kernel Methods in R. Journal of Statistical Software, 11(9), 1-20. URL http://www.jstatsoft.org/v11/i09/

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<[doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)>.

Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. [https://CRAN.R-project.org/package=tidyverse](https://cran.r-project.org/package=tidyverse)

Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read - Rectangular Text Data. R package version 1.1.1. [https://CRAN.R-project.org/package=readr](https://cran.r-project.org/package=readr)

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Max Kuhn and Hadley Wickham (2019). tidymodels: Easily Install and Load the 'Tidymodels' Packages. R package version 0.0.3.

[https://CRAN.R-project.org/package=tidymodels](https://cran.r-project.org/package=tidymodels)

Max Kuhn and Davis Vaughan (2019). parsnip: A Common API to Modeling and Analysis Functions. R package version 0.0.3.1. [https://CRAN.R-project.org/package=parsnip](https://cran.r-project.org/package=parsnip)

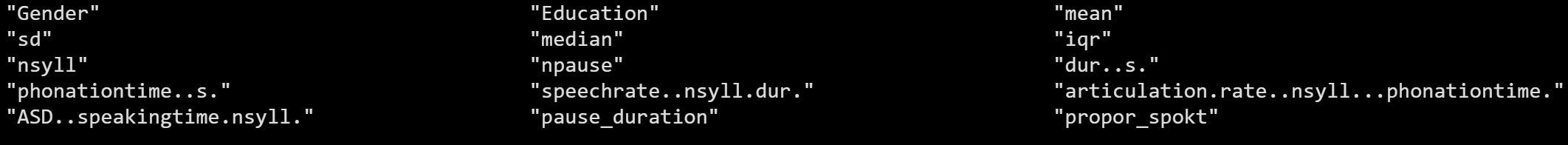
Max Kuhn and Ross Quinlan (2018). C50: C5.0 Decision Trees and Rule-Based Models. R package version 0.1.2. [https://CRAN.R-project.org/package=C50](https://cran.r-project.org/package=C50)

Ludvig Renbo Olsen (2019). groupdata2: Creating Groups from Data. R package version 1.1.2. https://CRAN.R-project.org/package=groupdata2

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL: <http://www.rstudio.com/>.

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, [https://CRAN.R-project.org/package=psych](https://cran.r-project.org/package=psych) Version = 1.8.12.

### Appendix

**