Portfolio 3

Experimental Methods 3: Multilevel models and machine learning

E22.147201U013.A

Submission: 28. 11 2022

Student: Sara Kjær Kristensen (au702259@uni.au.dk)

Maria Mujemula Olsen (au695881@uni.au.dk)

School of Communication and Culture, University of Aarhus

Nordre Ringgade 104 4 th, 8200 Aarhus N, Denmark

The assignment

The Machine Learning assignment has 3 main parts: First we create a skeptical and an informed simulation, based on the meta-analysis. Second, we build and test our machine learning pipeline on the simulated data. Third, we apply the pipeline to the empirical data.

The report for the exam, thus, consists of the answer to all the following prompts:

* Describe your machine learning pipeline. Produce a diagram of it to guide the reader (e.g. see Rybner et al 2022 Vocal markers of autism: Assessing the generalizability of ML models), and describe the different parts: data budgeting, data preprocessing, model choice and training, assessment of performance.
* Briefly justify and describe your use of simulated data, and results from the pipeline on them.
* Describe results from applying the ML pipeline to the empirical data and what can we learn from them.

Remember: plots are very very important to communicate your process and results.

### Part I - Simulating data

Use the meta-analysis reported in Parola et al., 2020, create a simulated dataset with 100 matched pairs of schizophrenia and controls, each participant producing 10 repeated measures (10 trials with their speech recorded). for each of these "recordings" (data points) produce 10 acoustic measures: 6 from the meta-analysis, 4 with just random noise. Do the same for a baseline dataset including only 10 noise variables. Tip: see the slides for the code.

Using the meta-analysis of voice patterns in schizophrenic (SC) and control (C), we simulate two datasets with 100 pairs of SC and C through 10 trials. We make two dataset one with informed and one with sceptic estimates of pitch mode (PitchMode), pitch variability (PitchVar), proportion of spoken time (ProSpotime), speech rate (SpeechRate), number (NumPause) and length (LenPause) of pauses, and an additional 4 noise measurements (see figure 1).

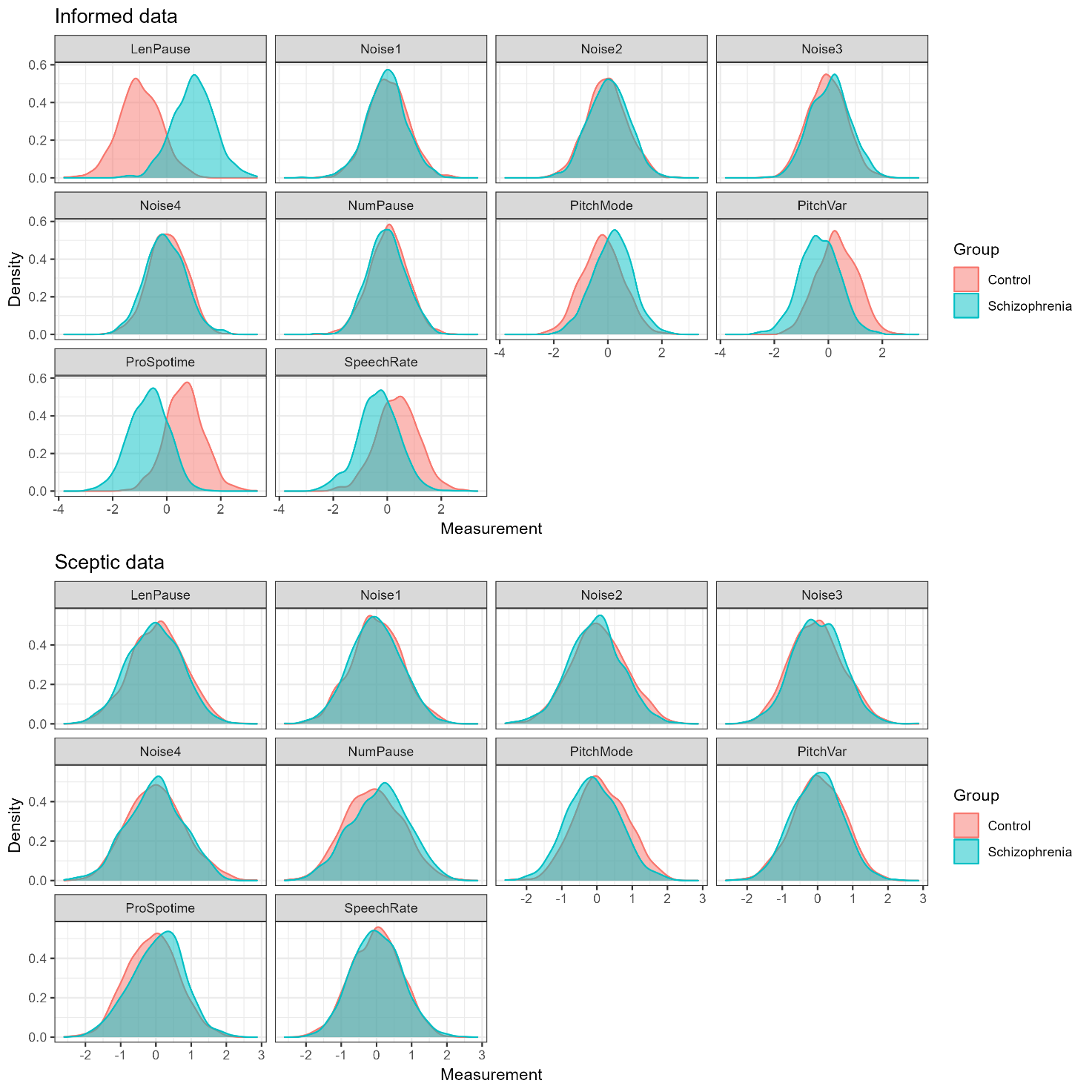


Figure 1 - Simulated data with comparing participant group estimates for informed and sceptic datasets.

### Part II - ML pipeline on simulated data

On the two simulated datasets (separately) build a machine learning pipeline: create a data budget (e.g., balanced training and test sets); pre-process the data (e.g., scaling the features); fit and assess a classification algorithm on the training data (e.g., Bayesian multilevel logistic regression); discuss whether performance is as expected, and feature importance is as expected.

Bonus question: replace the Bayesian multilevel regression with a different algorithm, e.g., SVM or random forest (but really, anything you'd like to try).

Besides testing our informed and sceptic datasets predictions, we test three different models using R (version 4.2.1) and the packages brms, cmdstanr and tidyverse (see references) for a bayesian approach: a baseline model with fixed effects, a model with varying intercept per participant, and lastly a model with varying intercepts for participants and slopes per measure.

We start of by splitting our data with an 80/20 ratio for training and test sets for both datasets. Next, we standardize our now four separate sets measures. We fix normally distributed priors per class; mean of 0 and standard deviation of 1 for all measures and intercepts, mean of 0 and standard deviation of 0.1 for the measure’s standard deviations, and a standard Lewandowski, Kurowicka and Joe correlation of 1. After fitting our priors and training to our models using Bernoulli, we asses prior posterior update plots for all relevant variables. The common green colour is for posterior and the common red colour is for the prior distribution.

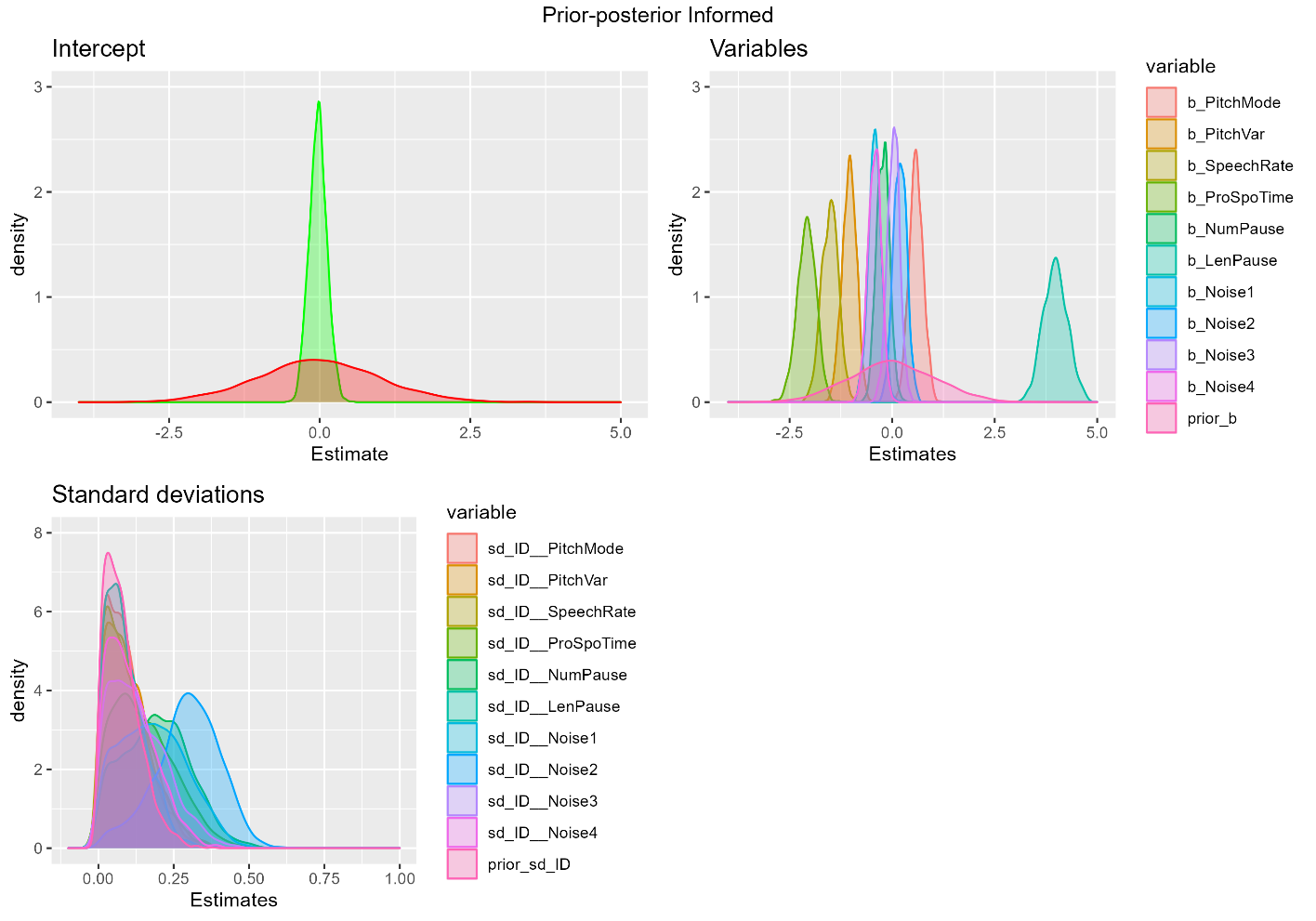


Figure 2 - Informed training dataset fitted to model with varying intercepts and slopes

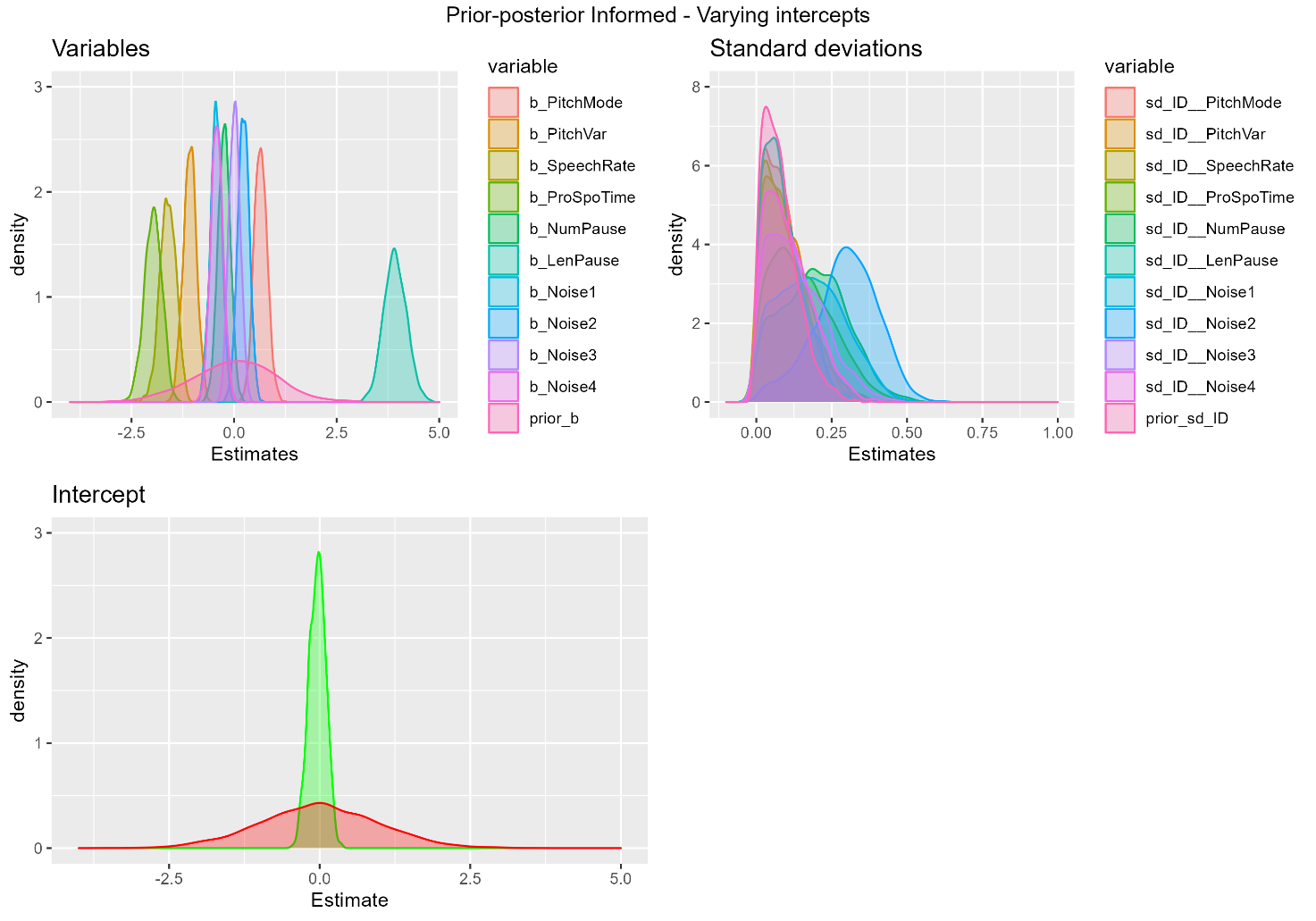


Figure 3 - Informed training dataset fitted to model with varying intercepts

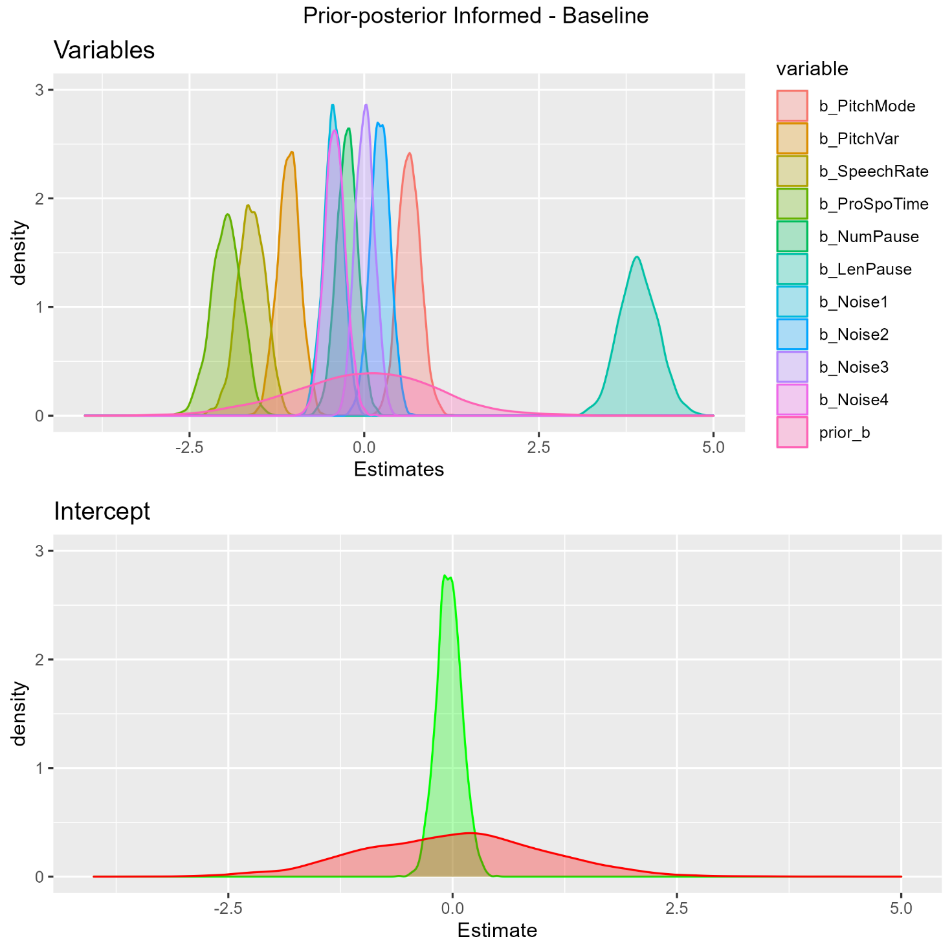


Figure 4 - Informed training dataset fitted to baseline model

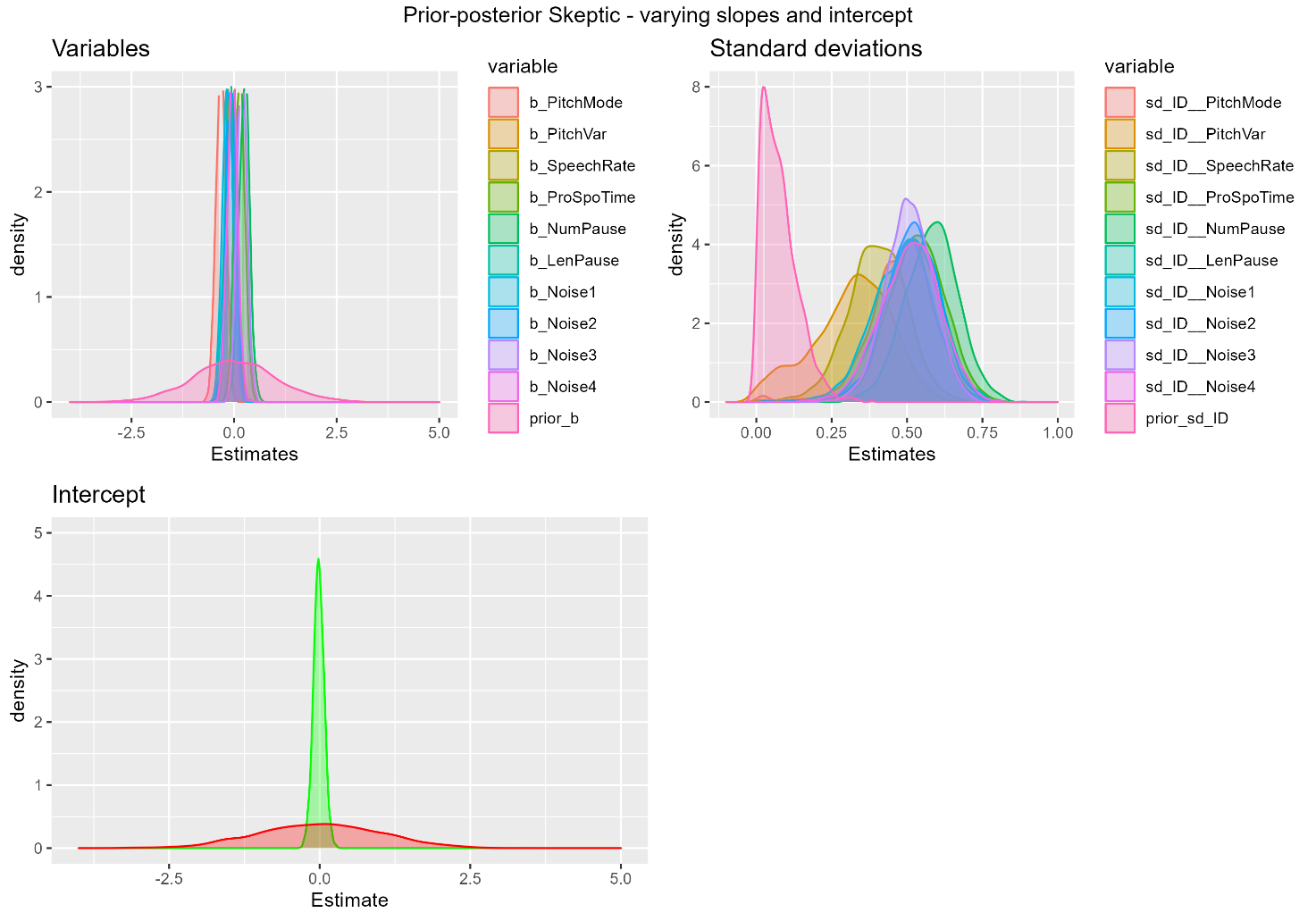


Figure 5 - Sceptic training dataset fitted to model with varying slopes and intercepts

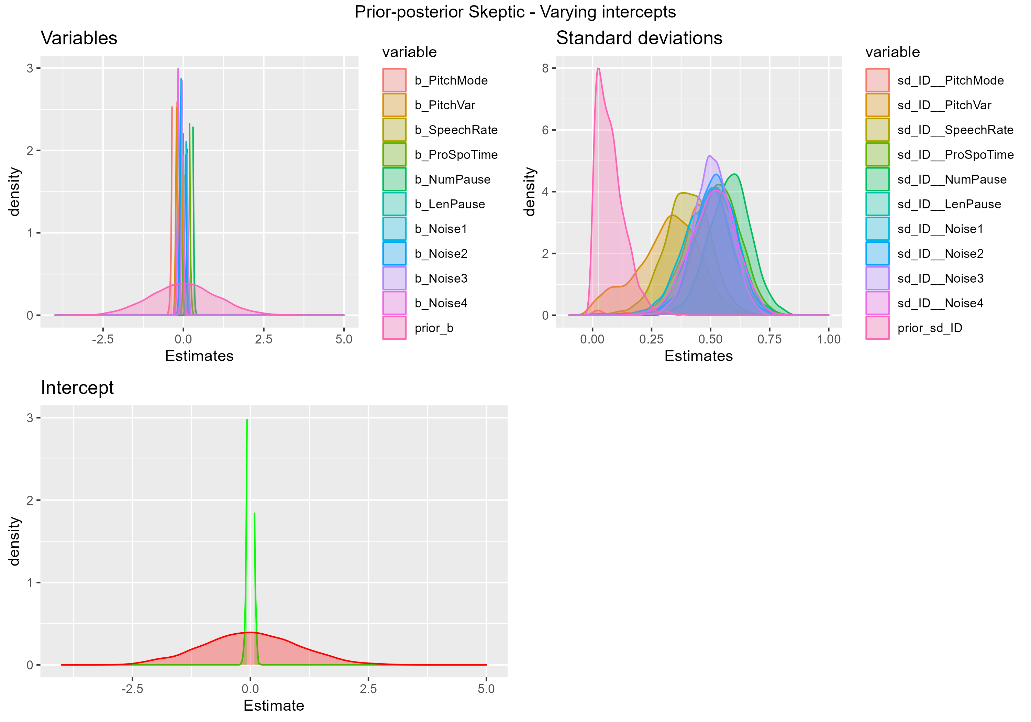


Figure 6 - sceptic training dataset fitted to model with varying intercepts

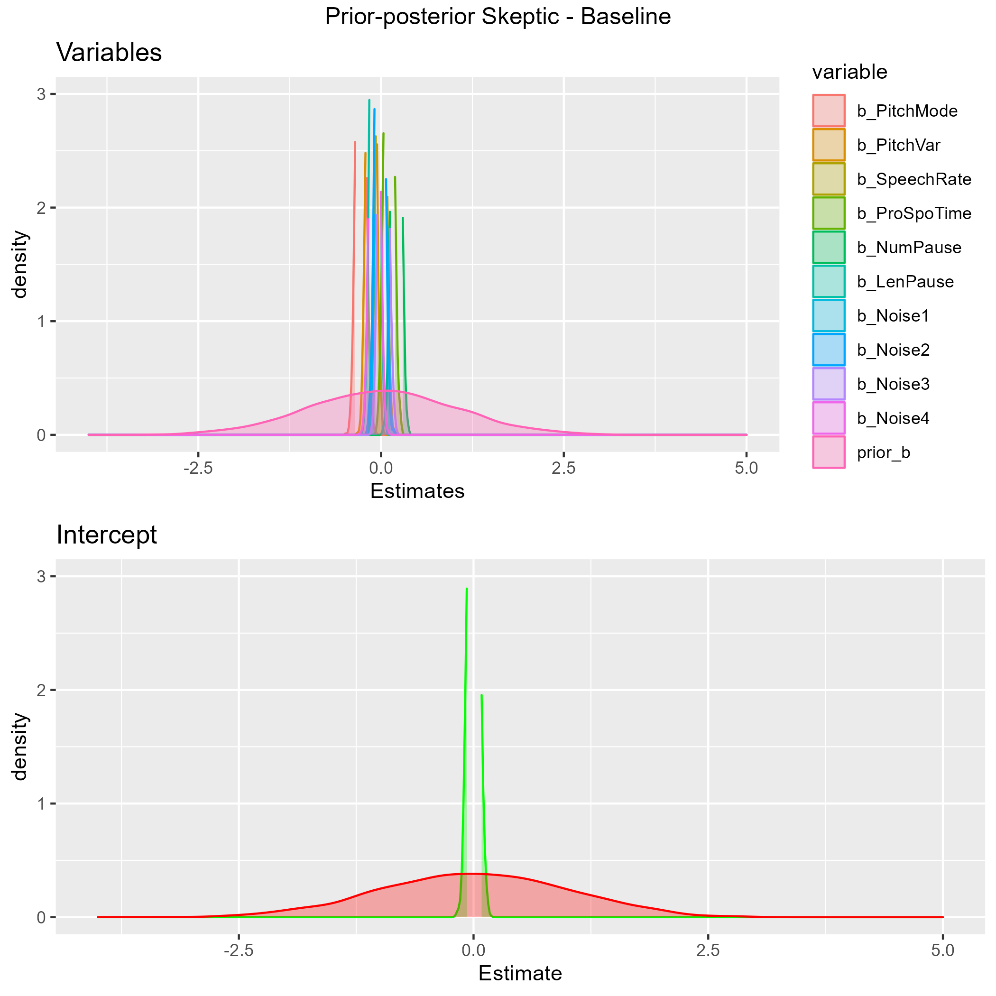


Figure 7 - Sceptic training set fitted to baseline model

Do we need a sensitivity analysis of the prior’s influence?

We now compare the different models using leave one out weighted method.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Informed | | | | Sceptic | | | |
| Elpd\_diff | Se\_diff | Weight | Model | | Elpd\_diff | Se\_diff | Weight |
| 0.0 | 0.0 | 1.0 | Varying slopes and intercepts | | 0.0 | 0.0 | 1.0 |
| -32.3 | 4.0 | 0.0 | Varying intercepts | | -783.5 | 8.8 | 0.0 |
| -32.0 | 4.0 | 0.0 | Baseline | | -784.3 | 8.8 | 0.0 |

Well looking at this table, the varying slopes and intercept model seems to be performing the best, then the baseline model and lastly the varying intercepts. But as all the information possible is in the first model, it seems rather ambiguous to say that we cannot learn as much from the other models. Therefor we check their predictive performance on their respectable test sets.

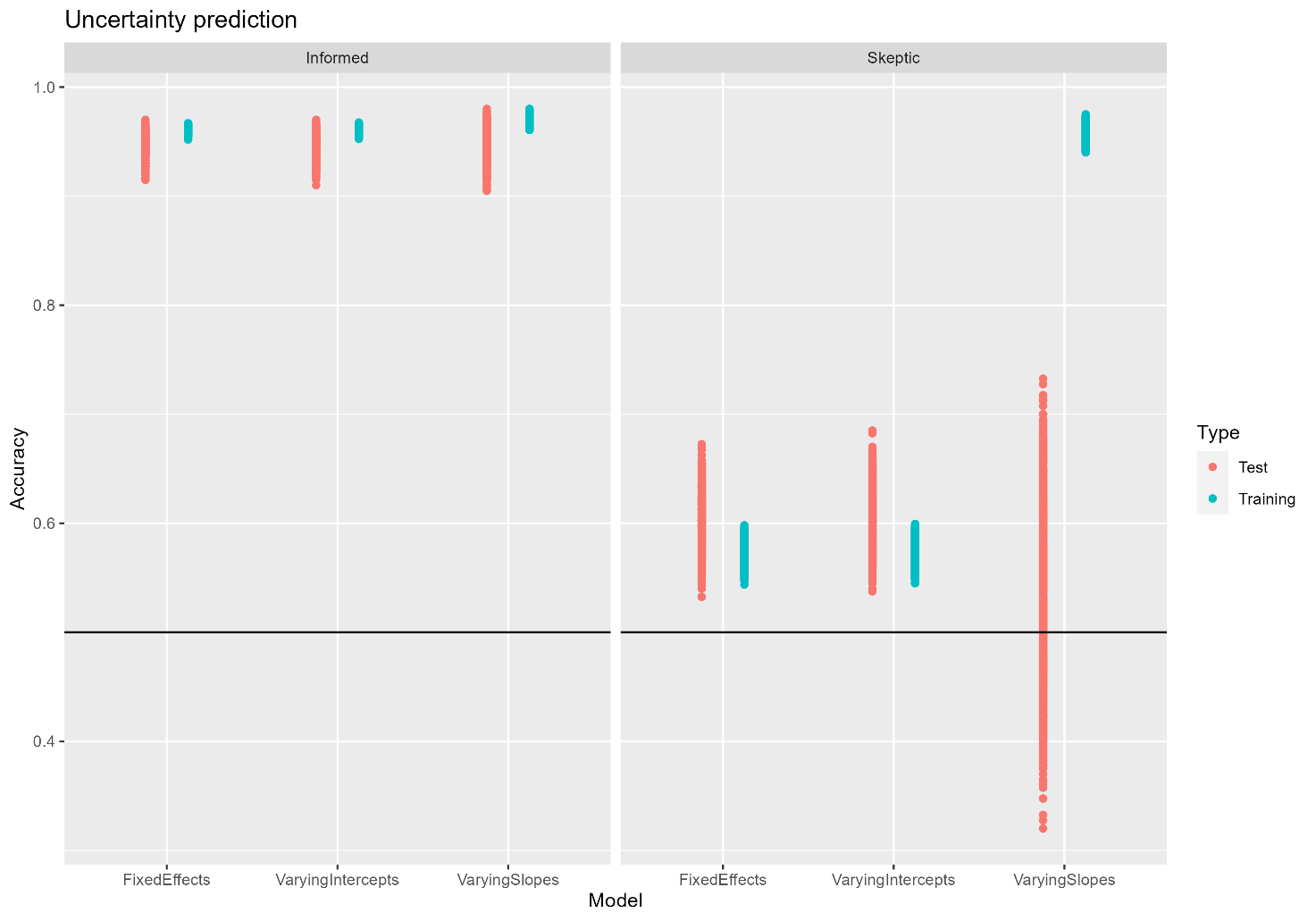


Figure 8 - Prediction with uncertainty

Here the models are in the order of baseline, varying intercepts, varying slopes and intercepts. The plot paints a rather clear picture of the varying slopes model over fitting likewise the informed models. We may for future reference use a sceptic set of priors and baseline or varying intercept model.

Looking at the models output:

|  |  |  |
| --- | --- | --- |
| Informed | Model | Sceptic |
| Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  sd(Intercept) 0.07 0.06 0.00 0.20 1.00 1363 876  sd(PitchMode) 0.09 0.07 0.00 0.26 1.00 1057 906  sd(PitchVar) 0.10 0.08 0.01 0.28 1.00 1136 797  sd(SpeechRate) 0.10 0.07 0.00 0.27 1.00 1061 854  sd(ProSpoTime) 0.15 0.10 0.01 0.37 1.00 790 864  sd(NumPause) 0.19 0.11 0.01 0.41 1.00 649 643  sd(LenPause) 0.09 0.06 0.01 0.24 1.00 1820 1284  sd(Noise1) 0.18 0.11 0.01 0.40 1.00 595 672  sd(Noise2) 0.30 0.11 0.06 0.48 1.00 811 410  sd(Noise3) 0.13 0.09 0.01 0.34 1.00 770 758  sd(Noise4) 0.11 0.08 0.01 0.29 1.00 955 749  cor(Intercept,PitchMode) 0.02 0.29 -0.54 0.57 1.00 2342 1529  cor(Intercept,PitchVar) -0.01 0.29 -0.57 0.57 1.00 3198 1330  cor(PitchMode,PitchVar) 0.01 0.29 -0.54 0.56 1.00 2112 1559  cor(Intercept,SpeechRate) 0.01 0.29 -0.54 0.57 1.00 3051 1632  cor(PitchMode,SpeechRate) 0.01 0.29 -0.54 0.55 1.00 2034 1419  cor(PitchVar,SpeechRate) -0.01 0.29 -0.55 0.54 1.00 1691 1448  cor(Intercept,ProSpoTime) -0.01 0.29 -0.55 0.55 1.00 1700 1368  cor(PitchMode,ProSpoTime) -0.00 0.28 -0.53 0.52 1.00 2124 1334  cor(PitchVar,ProSpoTime) 0.03 0.29 -0.54 0.58 1.00 1395 1329  cor(SpeechRate,ProSpoTime) -0.04 0.29 -0.57 0.52 1.00 1474 1417  cor(Intercept,NumPause) 0.01 0.29 -0.55 0.55 1.00 1360 1469  cor(PitchMode,NumPause) 0.01 0.30 -0.53 0.59 1.00 1464 1244  cor(PitchVar,NumPause) -0.04 0.29 -0.59 0.51 1.00 1111 1200  cor(SpeechRate,NumPause) 0.06 0.28 -0.49 0.59 1.00 1350 1472  cor(ProSpoTime,NumPause) -0.11 0.29 -0.62 0.47 1.00 1087 1283  cor(Intercept,LenPause) 0.00 0.29 -0.55 0.54 1.00 2762 1252  cor(PitchMode,LenPause) 0.00 0.30 -0.55 0.59 1.00 2279 1504  cor(PitchVar,LenPause) -0.01 0.29 -0.57 0.53 1.00 2071 1721  cor(SpeechRate,LenPause) 0.02 0.28 -0.53 0.56 1.00 1850 1378  cor(ProSpoTime,LenPause) -0.03 0.28 -0.56 0.52 1.00 1795 1529  cor(NumPause,LenPause) 0.03 0.29 -0.53 0.60 1.00 1705 1828  cor(Intercept,Noise1) -0.01 0.29 -0.55 0.54 1.00 1748 1395  cor(PitchMode,Noise1) 0.00 0.29 -0.55 0.58 1.00 1647 1321  cor(PitchVar,Noise1) 0.03 0.29 -0.52 0.57 1.00 1494 1403  cor(SpeechRate,Noise1) -0.03 0.29 -0.58 0.54 1.00 1730 1552  cor(ProSpoTime,Noise1) 0.10 0.29 -0.48 0.62 1.00 1228 1382  cor(NumPause,Noise1) -0.12 0.29 -0.62 0.46 1.00 1049 1682  cor(LenPause,Noise1) -0.03 0.29 -0.57 0.54 1.00 1013 1493  cor(Intercept,Noise2) -0.00 0.29 -0.55 0.56 1.00 882 1263  cor(PitchMode,Noise2) -0.03 0.29 -0.57 0.53 1.00 954 1303  cor(PitchVar,Noise2) 0.06 0.28 -0.48 0.58 1.00 927 1426  cor(SpeechRate,Noise2) -0.08 0.28 -0.60 0.46 1.00 1289 1524  cor(ProSpoTime,Noise2) 0.17 0.28 -0.42 0.66 1.00 955 1337  cor(NumPause,Noise2) -0.29 0.27 -0.73 0.34 1.00 822 1226  cor(LenPause,Noise2) -0.06 0.29 -0.59 0.51 1.00 1664 1732  cor(Noise1,Noise2) 0.23 0.27 -0.36 0.70 1.00 1031 1267  cor(Intercept,Noise3) -0.00 0.29 -0.53 0.54 1.00 2106 1708  cor(PitchMode,Noise3) 0.02 0.28 -0.55 0.54 1.00 1800 1376  cor(PitchVar,Noise3) 0.01 0.29 -0.54 0.57 1.00 1819 1496  cor(SpeechRate,Noise3) 0.03 0.29 -0.53 0.57 1.00 1618 1425  cor(ProSpoTime,Noise3) -0.02 0.29 -0.57 0.55 1.00 1414 1427  cor(NumPause,Noise3) 0.03 0.29 -0.52 0.58 1.00 1534 1585  cor(LenPause,Noise3) -0.01 0.29 -0.58 0.56 1.00 1403 1432  cor(Noise1,Noise3) 0.02 0.28 -0.53 0.53 1.00 1735 1528  cor(Noise2,Noise3) -0.02 0.28 -0.56 0.50 1.00 1395 1891  cor(Intercept,Noise4) -0.01 0.28 -0.54 0.55 1.00 2225 1762  cor(PitchMode,Noise4) 0.02 0.29 -0.53 0.58 1.00 2327 1356  cor(PitchVar,Noise4) -0.01 0.29 -0.55 0.55 1.00 2039 1596  cor(SpeechRate,Noise4) -0.01 0.29 -0.56 0.56 1.00 1878 1667  cor(ProSpoTime,Noise4) -0.01 0.28 -0.55 0.54 1.00 1306 1474  cor(NumPause,Noise4) 0.03 0.28 -0.51 0.56 1.00 1414 1668  cor(LenPause,Noise4) 0.01 0.28 -0.52 0.56 1.00 1494 1446  cor(Noise1,Noise4) 0.02 0.29 -0.54 0.57 1.00 1601 1673  cor(Noise2,Noise4) -0.04 0.29 -0.56 0.54 1.00 1662 1654  cor(Noise3,Noise4) -0.01 0.29 -0.56 0.54 1.00 1205 1573  Population-Level Effects:  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  Intercept -0.03 0.15 -0.30 0.27 1.00 2482 1244  PitchMode 0.59 0.17 0.26 0.93 1.00 2193 1336  PitchVar -1.04 0.17 -1.39 -0.71 1.00 2035 1655  SpeechRate -1.52 0.20 -1.92 -1.12 1.00 2128 1655  ProSpoTime -2.08 0.23 -2.55 -1.65 1.00 2008 1416  NumPause -0.22 0.16 -0.52 0.08 1.00 2277 1525  LenPause 3.98 0.29 3.42 4.57 1.00 1472 1558  Noise1 -0.42 0.16 -0.75 -0.10 1.00 1971 1342  Noise2 0.19 0.16 -0.12 0.50 1.00 2462 1641  Noise3 0.04 0.15 -0.27 0.33 1.00 2169 1191  Noise4 -0.43 0.16 -0.74 -0.13 1.00 2527 1551 | Diagnosis ~ 1 + PitchVar + SpeechRate + ProSpoTime + NumPause + LenPause + Noise1 + Noise2 + Noise3 + Noise4 + (1 + PitchVar + SpeechRate + ProSpoTime + NumPause + LenPause + Noise1 + Noise2 + Noise3 + Noise4 | ID) | Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  sd(Intercept) 0.05 0.04 0.00 0.14 1.00 1983 1358  sd(PitchMode) 0.48 0.11 0.25 0.67 1.00 718 365  sd(PitchVar) 0.32 0.13 0.03 0.54 1.00 469 400  sd(SpeechRate) 0.40 0.10 0.19 0.58 1.00 1028 827  sd(ProSpoTime) 0.53 0.09 0.34 0.71 1.00 1531 1079  sd(NumPause) 0.58 0.09 0.41 0.75 1.00 1371 1457  sd(LenPause) 0.51 0.10 0.32 0.69 1.00 1308 1353  sd(Noise1) 0.48 0.10 0.28 0.66 1.00 1047 888  sd(Noise2) 0.51 0.09 0.32 0.68 1.00 1080 632  sd(Noise3) 0.51 0.08 0.35 0.66 1.00 2061 1700  sd(Noise4) 0.51 0.10 0.31 0.69 1.00 1209 1186  cor(Intercept,PitchMode) -0.02 0.29 -0.58 0.52 1.00 405 934  cor(Intercept,PitchVar) -0.01 0.28 -0.55 0.52 1.00 1206 1521  cor(PitchMode,PitchVar) -0.12 0.24 -0.56 0.39 1.00 1655 1639  cor(Intercept,SpeechRate) 0.01 0.29 -0.55 0.58 1.00 547 1181  cor(PitchMode,SpeechRate) -0.10 0.21 -0.49 0.31 1.00 2097 1652  cor(PitchVar,SpeechRate) -0.05 0.25 -0.53 0.46 1.00 1135 1046  cor(Intercept,ProSpoTime) 0.02 0.29 -0.55 0.56 1.00 429 687  cor(PitchMode,ProSpoTime) -0.31 0.18 -0.67 0.05 1.00 1336 1534  cor(PitchVar,ProSpoTime) -0.04 0.24 -0.49 0.45 1.00 800 1301  cor(SpeechRate,ProSpoTime) 0.17 0.20 -0.25 0.56 1.00 1218 1319  cor(Intercept,NumPause) -0.01 0.27 -0.53 0.50 1.00 538 1026  cor(PitchMode,NumPause) 0.26 0.19 -0.13 0.61 1.00 1111 1487  cor(PitchVar,NumPause) 0.05 0.22 -0.37 0.49 1.00 1067 1146  cor(SpeechRate,NumPause) 0.00 0.20 -0.39 0.38 1.00 1339 1640  cor(ProSpoTime,NumPause) -0.13 0.18 -0.48 0.21 1.00 1279 1441  cor(Intercept,LenPause) 0.00 0.29 -0.55 0.52 1.00 587 906  cor(PitchMode,LenPause) -0.16 0.19 -0.53 0.21 1.00 1250 1381  cor(PitchVar,LenPause) -0.20 0.24 -0.62 0.27 1.00 794 1111  cor(SpeechRate,LenPause) 0.23 0.20 -0.19 0.60 1.00 1124 1340  cor(ProSpoTime,LenPause) -0.17 0.19 -0.52 0.21 1.00 1730 1962  cor(NumPause,LenPause) 0.15 0.18 -0.23 0.49 1.00 2004 1698  cor(Intercept,Noise1) 0.00 0.30 -0.56 0.57 1.00 563 1216  cor(PitchMode,Noise1) 0.07 0.20 -0.33 0.46 1.00 1187 1584  cor(PitchVar,Noise1) 0.14 0.24 -0.34 0.58 1.00 899 1101  cor(SpeechRate,Noise1) 0.26 0.20 -0.15 0.62 1.00 1151 1412  cor(ProSpoTime,Noise1) -0.10 0.19 -0.46 0.29 1.00 1441 1898  cor(NumPause,Noise1) 0.08 0.19 -0.30 0.44 1.00 2054 1619  cor(LenPause,Noise1) -0.03 0.20 -0.42 0.36 1.00 1612 1647  cor(Intercept,Noise2) 0.00 0.29 -0.54 0.55 1.00 467 1053  cor(PitchMode,Noise2) 0.14 0.19 -0.24 0.52 1.00 1143 1298  cor(PitchVar,Noise2) 0.17 0.23 -0.32 0.59 1.00 836 999  cor(SpeechRate,Noise2) 0.27 0.20 -0.14 0.63 1.00 1162 1576  cor(ProSpoTime,Noise2) -0.25 0.19 -0.60 0.12 1.00 1702 1672  cor(NumPause,Noise2) -0.03 0.18 -0.39 0.34 1.00 2029 1925  cor(LenPause,Noise2) 0.07 0.20 -0.31 0.45 1.00 1546 1751  cor(Noise1,Noise2) 0.16 0.19 -0.22 0.52 1.00 1894 1465  cor(Intercept,Noise3) 0.02 0.29 -0.56 0.57 1.00 495 674  cor(PitchMode,Noise3) -0.28 0.19 -0.61 0.12 1.00 1299 1245  cor(PitchVar,Noise3) 0.01 0.23 -0.41 0.46 1.00 1144 1505  cor(SpeechRate,Noise3) 0.35 0.20 -0.06 0.72 1.00 1343 1631  cor(ProSpoTime,Noise3) 0.43 0.17 0.07 0.74 1.00 1654 1561  cor(NumPause,Noise3) 0.13 0.18 -0.23 0.47 1.00 1967 1634  cor(LenPause,Noise3) 0.08 0.19 -0.30 0.46 1.00 1951 1965  cor(Noise1,Noise3) -0.10 0.19 -0.46 0.28 1.00 1812 1788  cor(Noise2,Noise3) 0.07 0.19 -0.32 0.40 1.00 2093 1613  cor(Intercept,Noise4) -0.00 0.29 -0.55 0.57 1.00 476 763  cor(PitchMode,Noise4) 0.14 0.20 -0.24 0.54 1.00 1202 1352  cor(PitchVar,Noise4) 0.09 0.24 -0.40 0.56 1.00 795 1134  cor(SpeechRate,Noise4) 0.15 0.21 -0.27 0.54 1.00 1286 1417  cor(ProSpoTime,Noise4) -0.07 0.19 -0.42 0.30 1.00 1502 1632  cor(NumPause,Noise4) 0.21 0.18 -0.14 0.54 1.00 1975 1693  cor(LenPause,Noise4) -0.14 0.20 -0.52 0.27 1.00 1592 1638  cor(Noise1,Noise4) 0.26 0.20 -0.15 0.63 1.00 1599 1753  cor(Noise2,Noise4) 0.00 0.20 -0.40 0.41 1.00 1567 1687  cor(Noise3,Noise4) 0.09 0.18 -0.28 0.45 1.00 1564 1812  Population-Level Effects:  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  Intercept -0.02 0.09 -0.19 0.16 1.00 4009 1504  PitchMode -0.32 0.12 -0.56 -0.10 1.00 2845 1517  PitchVar -0.07 0.11 -0.30 0.16 1.00 3833 1272  SpeechRate -0.05 0.11 -0.27 0.18 1.00 3205 1607  ProSpoTime 0.14 0.12 -0.10 0.38 1.00 2569 1821  NumPause 0.25 0.13 0.00 0.50 1.00 3606 1485  LenPause -0.11 0.13 -0.37 0.13 1.00 2869 1633  Noise1 -0.16 0.12 -0.40 0.08 1.00 3418 1364  Noise2 -0.08 0.12 -0.31 0.15 1.00 3084 1662  Noise3 0.19 0.12 -0.04 0.43 1.00 3315 1381  Noise4 -0.04 0.12 -0.28 0.21 1.00 2923 1394 |
| Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  sd(Intercept) 0.07 0.06 0.00 0.20 1.00 1249 1112  Population-Level Effects:  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  Intercept -0.04 0.14 -0.31 0.22 1.00 3244 1569  PitchMode 0.65 0.16 0.34 0.97 1.00 2361 1160  PitchVar -1.08 0.16 -1.40 -0.77 1.00 2108 1472  SpeechRate -1.62 0.20 -2.03 -1.25 1.00 1967 1343  ProSpoTime -1.98 0.21 -2.41 -1.59 1.00 2388 1482  NumPause -0.24 0.15 -0.54 0.04 1.00 2384 1473  LenPause 3.93 0.28 3.40 4.48 1.00 1800 1556  Noise1 -0.44 0.14 -0.73 -0.17 1.00 2417 1404  Noise2 0.23 0.14 -0.05 0.50 1.00 2878 1703  Noise3 0.01 0.13 -0.24 0.28 1.00 2845 1531  Noise4 -0.42 0.15 -0.71 -0.13 1.00 1843 1228 | Diagnosis ~ 1 + PitchMode + PitchVar + SpeechRate + ProSpoTime + NumPause + LenPause + Noise1 + Noise2 + Noise3 + Noise4 + (1 | ID) | Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  sd(Intercept) 0.04 0.03 0.00 0.10 1.00 1849 1063  Population-Level Effects:  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  Intercept 0.00 0.05 -0.10 0.10 1.00 2658 1356  PitchMode -0.28 0.05 -0.38 -0.17 1.00 2757 1619  PitchVar -0.14 0.05 -0.24 -0.04 1.00 3615 1428  SpeechRate 0.00 0.05 -0.10 0.10 1.00 3101 1648  ProSpoTime 0.11 0.05 0.01 0.21 1.00 2861 1496  NumPause 0.21 0.05 0.10 0.32 1.00 4368 1457  LenPause -0.09 0.05 -0.19 0.01 1.00 3583 1290  Noise1 -0.00 0.05 -0.11 0.11 1.00 4187 1220  Noise2 -0.02 0.05 -0.12 0.09 1.00 2924 1640  Noise3 0.04 0.05 -0.06 0.14 1.00 3055 1120  Noise4 -0.09 0.05 -0.19 0.00 1.00 3937 1502 |
| Population-Level Effects:  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  Intercept -0.04 0.14 -0.31 0.24 1.00 2242 1392  PitchMode 0.65 0.15 0.34 0.95 1.00 2870 1648  PitchVar -1.08 0.17 -1.42 -0.77 1.00 2757 1661  SpeechRate -1.61 0.19 -2.01 -1.24 1.00 2284 1611  ProSpoTime -1.98 0.21 -2.39 -1.59 1.00 2349 1242  NumPause -0.24 0.14 -0.53 0.03 1.00 3214 1569  LenPause 3.93 0.28 3.39 4.49 1.00 2394 1432  Noise1 -0.45 0.14 -0.72 -0.19 1.00 2863 1751  Noise2 0.23 0.14 -0.04 0.50 1.00 2483 1637  Noise3 0.01 0.14 -0.27 0.28 1.00 2599 1231  Noise4 -0.42 0.14 -0.70 -0.13 1.00 2883 1538 | Diagnosis ~ 1 + PitchMode + PitchVar + SpeechRate + ProSpoTime + NumPause + LenPause + Noise1 + Noise2 + Noise3 + Noise4 | Population-Level Effects:  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  Intercept 0.00 0.05 -0.10 0.10 1.00 2531 1484  PitchMode -0.28 0.05 -0.38 -0.17 1.00 2242 1545  PitchVar -0.13 0.05 -0.24 -0.03 1.00 2174 1359  SpeechRate 0.00 0.05 -0.10 0.10 1.00 2339 1259  ProSpoTime 0.11 0.05 0.00 0.21 1.00 1541 978  NumPause 0.21 0.05 0.11 0.31 1.00 1943 1660  LenPause -0.09 0.05 -0.18 0.01 1.00 2126 1662  Noise1 -0.00 0.05 -0.10 0.10 1.01 1937 1414  Noise2 -0.02 0.05 -0.12 0.09 1.00 2115 1659  Noise3 0.03 0.05 -0.07 0.14 1.00 1752 1292  Noise4 -0.09 0.05 -0.19 0.02 1.00 2247 1441 |

All models seem to be estimating probably. The informed varying intercepts and slopes model seem to be overfitting. For further investigations in voice patterns for schizophrenic, we will use the varying intercept models to allow participants to have their own starting point, but not fitting too neatly to the datapoints.

How should we include the model output?

### Part III - Applying the ML pipeline to empirical data

Download the empirical dataset from brightspace and apply your ML pipeline to the new data, adjusting where needed. Warning: in the simulated dataset we only had 10 features, now you have many more! Such is the life of the ML practitioner. Consider the impact a higher number of features will have on your ML inference, and decide whether you need to cut down the number of features before running the pipeline (or alternatively expand the pipeline to add feature selection).

Data: <https://www.dropbox.com/s/7ky1axvea33lgye/Ass3_empiricalData1.csv?dl=0>

Looking at the empirical data from Parola et al. (2020) we have 1889 observations over 398 variables. In total 122 participants with an equal number of females and males in the two diagnostic groups:

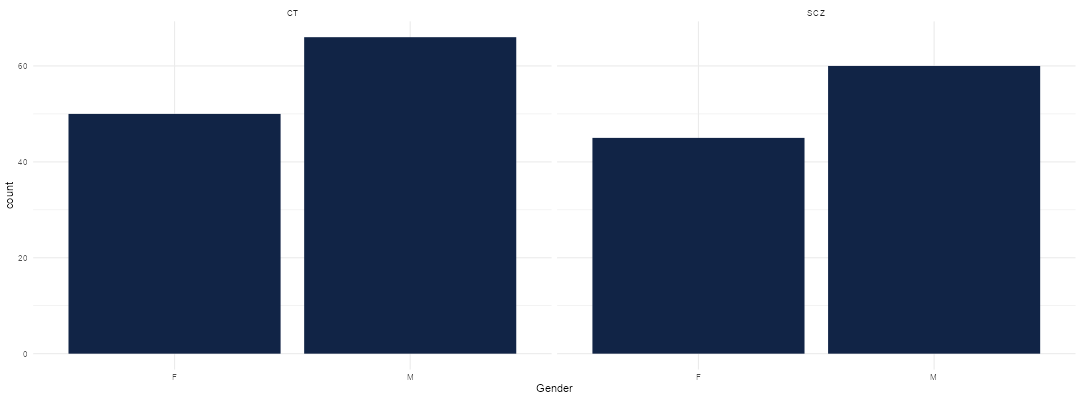


Figure 9 - Gender distribution in diagnostic groups in empirical data

Next plot displays a glimpse of diagnostic differences in duration of speech, number of pauses, percentage of spoken time and gender differences in mean pitch. For a more in-depth description of measurements see Parola et al. (2020).

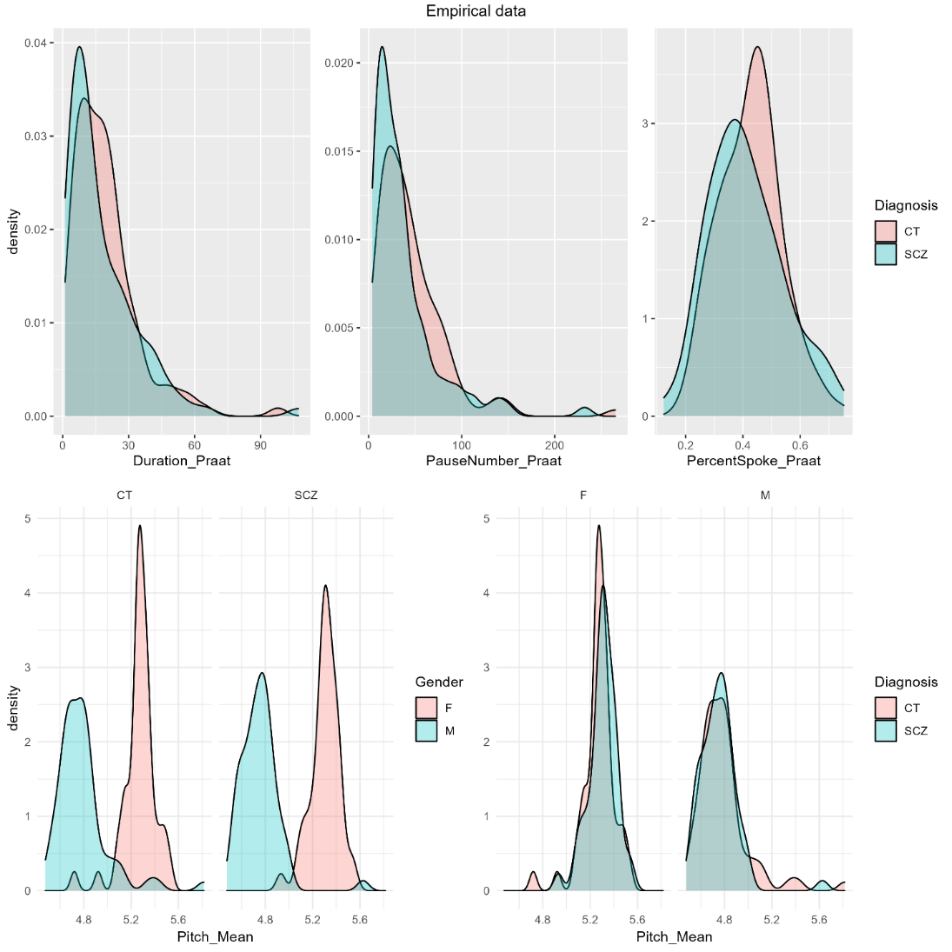
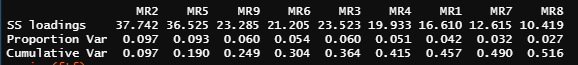


Figure 10 - A excerpt of diagnostic and gender differences in empirical data

Running a model on many highly correlated measures, can be done in a better way. Standardizing the measurements and applying a factor analysis, principal component analysis, we suggest using nine factors, to explore the dataset with. These nine factors will be the factor analysis scorer, looking at the table below one can get an overview of what the different scores contains, see appendix. Below each variables explained variance in the data set. The grand total of only 52% of the variance explained:



Using ID, diagnosis and the FA scores in a balanced training set, we fit an investigate prior posterior distribution of sceptic priors to our varying intercept model:

Diagnosis ~ MR2 + MR5 + MR9 + MR6 + MR3 + MR4 + MR1 + MR7 + MR8 + (1 | NewID)

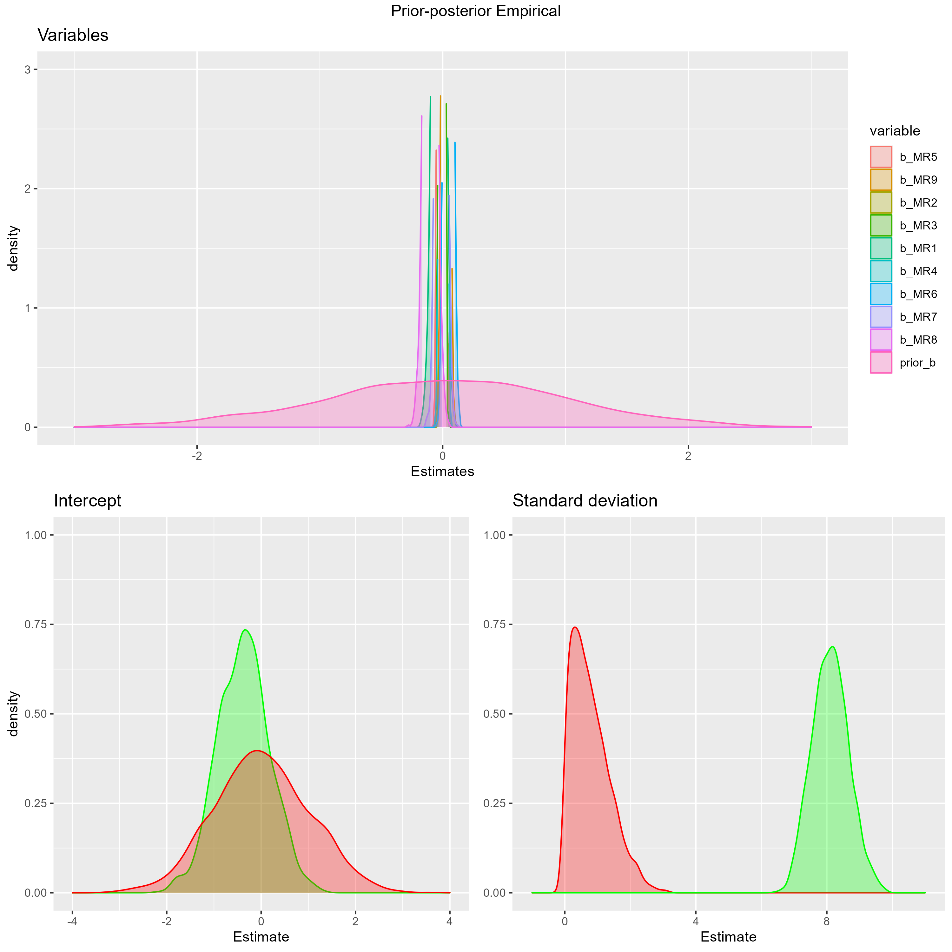


Figure 11 - Prior posterior distributions of empirical training set

No matter what prior we set for the standard deviation it kept having such a huge cap. We tried n(0,1) (above), n(0,2), n(0,4), n(0,6), it came slightly closer, but we could not see the idea of a standard deviation varying with that much. Would it make more sense to narrow the prior or what should we do?

Looking at the models output we get a group level effect of 8.11 [7.07-9.23] for the intercepts’ standard deviations. And several population level effects starting with the intercept with a probability of 75.65% of being -0.37 [-1.43:0.68] can not be considered significant nor unsignificant with a ROPE of 21.79%. See table for FA scores results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Estimate | CI | PD | ROPE |
| MR1 | -0.03 | -0.12 0.05 | 77.65% | 100% |
| MR2 | 0.01 | -1.43 0.68 | 83.75% | 100% |
| MR3 | -0.00 | -0.03 0.02 | 63.80% | 100% |
| MR4 | 0.01 | -0.03 0.04 | 63.95% | 100% |
| MR5 | -0.01 | -0.01 0.04 | 69.90% | 100% |
| MR6 | 0.05 | -0.00 0.10 | 96.85% | 100% |
| MR7 | -0.01 | -0.08 0.06 | 62.30% | 100% |
| MR8 | -0.10 | -0.18 -0.02 | 98.90% | 99.95% |
| MR9 | 0.03 | -0.01 0.07 | 88.55% | 100% |

Not sure how to report these findings/if we should have used a different method to sum/extract the variables from the 398 variables of interest.

The rest of the variables ROPE seems to point in the direction of negligible findings. All ROPE are in [-0.18 : 0.18]. Our models prediction accuracy, plotted below, seems to have a high fit.

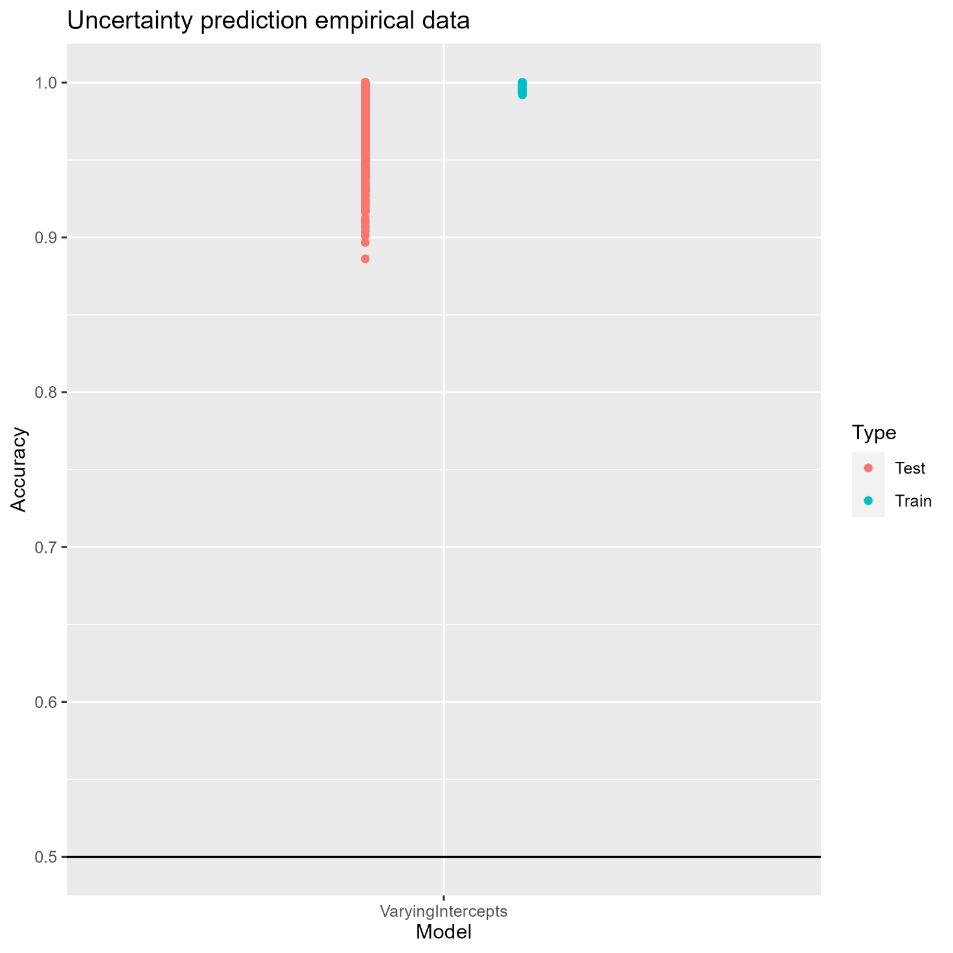


Figure 12 - Model prediction

Since our models’ variables are the collection of factor analysis scores, not much information is left out and our model is good at predicting the diagnosis outcome of our participants. However, it may be doing too good of a job. It would be nice to compare with another set of variables, e.g., one that has prioritized the most information rich variables in stead of keeping everything.

### References

Parola, A., Simonsen, A., Bliksted, V., & Fusaroli, R. (2020). Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. Schizophrenia Research, 216, 24–40. https://doi.org/10.1016/j.schres.2019.11.031

 R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel

Models Using Stan. Journal of Statistical Software, 80(1), 1-28.

doi:10.18637/jss.v080.i01

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund

G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM,

Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D,

Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” \_Journal of

Open Source Software\_, \*4\*(43), 1686. doi:10.21105/joss.01686

Gabry J, Češnovar R (2022). \_cmdstanr: R Interface to 'CmdStan'\_.

https://mc-stan.org/cmdstanr/, <https://discourse.mc-stan.org>.

### Appendix

