Normative modeling in Schizophrenia - Analysis of the 308 regions parcellation

Noemi González Lois

2021-06-02

Packages and libraries

```
rm(list = ls()) # clear the workspace
library(easypackages) # install.packages("easypackages")

# get a list of all the needed packages
list.of.packages <- c("viridis", "tidyverse", "MatchIt", "grid", "png",
                      "gridExtra", "parallel", "nlme", "JMbayes",
                      "BiocManager", "Biostrings", "lme4", "ggplot2",
                      "Hmisc", "devtools", "longCombat", "neuroCombat",
                      "tinytex")

# load them all
libraries(list.of.packages)
ncores = detectCores() # Number of cores available in pc

rm(list.of.packages)
```

Set working directory and load functions

Data preparation

Options of DataPreparation:

- parc = "parc35" or "parc308" (whether to use the 34 regions parcellation or the 308 regions one)
- harmonization= "lC" or "nC" (whether to use lonCombat or NeuroCombat harmonization)
- match = T or F (whether to use match-it or not)

Warning: Fewer control units than treated units; not all treated units will get a match.

Stack Overflow: This warning is because our treated group is larger than our control group (this happens in timepoint = 2). If you're doing 1:1 matching without replacement, all the control units will be used up before all the treated units get a match. **To remedy this, you need to match with replacement or think about whether you actually want to generalize to the control population and switch the labels on the treatment groups**. You can do this by creating a new variable, say notY, which is 1 - Y and then performing the same operations.

The match in *sex variable* is not exact. Anyway, the match in # patients, # controls is done well (¿is it enough?)

NO MATCHED DATASET

NO MATCHED	timepoint 1	timepoint 2	timepoint 3
# controls	298	293	109
# patients	169	168	50

Timepoint 1	sex 0	sex 1
# controls	131	167
# patients	38	131

Timepoint 2	sex 0	sex 1
# controls	130	163
# patients	38	130

Timepoint 3	sex 0	sex 1
# controls	50	59
# patients	7	43

MATCHED DATASET

Number of patients vs number of controls per timepoint is not exactly the same:

MATCHED	timepoint 1	timepoint 2	timepoint 3
# controls	169	164	49
# patients	169	164	49

Timepoint 1	sex 0	sex 1
# controls	38	131
# patients	38	131

Timepoint 2	sex 0	sex 1
# controls	37	127
# patients	38	126

Timepoint 3	sex 0	sex 1
# controls	8	41
# patients	7	42

Exploratory data analysis

Show relevant figures and analytics before and after data preparation. For example, age of controls vs age of patients in the raw df vs the match-it df:

Matching ages:

Linear Mixed Effects Model Regresion

Calling the RegressionModel function with different df. Return the z scores (one for each region for each subject)

Raw dataframe: