Data Analysis and Risk Assessment by

Using Longitudinal Data Analysis in Mental Disorder Diagnosis

Sunny Zeng, Ellen Lin

# Reference:

Long, J. (2012). Longitudinal data analysis for the behavioral sciences using R. Thousand Oaks, Calif.: SAGE.

# Goal:

Given a labeled dataset (i.e. TMI vs non-TMI), we analyze and distinguish the differences in the profiles’ behaviors, predict and forecast the trajectory model and its trends, and then assess risk factors among these groups. The problems we encounter are the followings:

* Feature Selection: What features and how many features to model?
* Use Case: given a recorded data or a set of meaningful signals at a time point (i.e. 9 months from the incidence), assess the risk of having the disorder.
* In the matter of fact, the model is sensitive to time, however there will be segmental changes over time of the same subject (i.e. a certain degree of elevation in some features over the past 3 months may “qualify” the subject to a higher risk.)
* Looking for influential features such as age and gender that help to assess risks (i.e. a boy is more likely to have autism than a girl)

# Methods and Approaches:

# Application of EEG data:

Result of the project:

* We found the trends and trajectory model of the EEG measurements for each groups: normal, high risk, and diagnosed with autism.
* given a set of measures over time from a new subject, we are able to compare against the group trajectories and determine whether the new trajectories are more likely to belong to the normal group or the autism group
* risk assessment given not just a one-time measurement but a trajectory of measurements

Data Structure:

* multiple records for one subject (at time points 3, 6, 9, 12, 18, 24, 36 months of age)
* Identifiers:
* ID: subject ID
* Age: time variable
* gender
* class: indicator (asd-autism, typ-normal, hra-high risk autism but do not have autism)
* Measurements:
* sensor or channel (19):
* ["C3", "C4", "O1", "O2", "Cz", "F3", "F4", "F7", "F8", "Fz", "Fp1", "Fp2", "P3", "P4", "Pz", "T7", "T8", "P7", "P8"]
* features (9):
* ['Power', 'SampE', 'RR', 'DET', 'LAM', 'L\_entr', 'L\_max', 'L\_mean', 'TT']
* Frequency band or scale (6):   [0,1,2,3,4,5]
* Example: C3.Power.s0

Key Steps

* plot profile plots for subjects from asd and typ group
* observe/fit trends
* develop a determination method
* Issues
* feature selection
* one more dimension to consider--time!
* some activities may only be caught by specific scale of the channel-feature pair
* too many features (1025)
* utilize time variable
* slopes between two time points
* fit trajectories with polynomials, make the coefficients as the new features for each subject of a given measurement
* Approaches
* feature selection:
* ordinary feature selection tools only depended on one-time measurement, do not consider time or change over time.
* we want to make our model sensitive to time or change over time.
* feature ranking based on differences on average trajectories of the groups
* for each feature, find the average trajectories for the groups
* there are differences in distances/area between the curves
* there are differences in the coefficients when fit the curves with polynomials
* slope comparison