Schizophrenia background and phenotype association

1 Background

Schizophrenia is a mental disorder characterized by abnormal social behavior and failure to understand what is real. An interaction between genes and a range of environment factors may cause schizophrenia. Neurotransmitters, dopamine and glutamate may affect schizophrenia. Schizophrenia typically begins in late adolescence or early adulthood. It is more common among males than females. It can be treated, about 20% of people do well, and a few recover completely. Schizophrenia often divide into positive and negative symptoms. Positive symptoms include hallucinations, delusions, thought disorders and movement disorders. It's usually can be cured by medicine to some extent. As for negative symptoms patients show social withdrawal, thinking volitional activity drops. Even schizophrenes take medicine, it's hard for them to recover.

We have 171 samples coming from community and Research center for mental disorders, includes 81 healthy controls (HC) and 90 schizophrenes (SZ). All of them do questionnaire and MATRICS Consensus Cognitive Battery (MCCB) test, but for patients, besides that, they need to do Positive and Negative Syndrome Scale (PANSS) test.

2 Scale

2.1 MATRICS Consensus Cognitive Battery (MCCB)

The MCCB is used to be an endpoint for clinical trials aiming to enhance cognition in schizophrenia [3]. It has 9 subtests, represents 7 aspects. After adjusted, higher scores mean higher levels of cognition.

In our study, both HC and SZ did this test, but MCCB scale contains about 40 NAs (see ../../SZData/mccb.txt).

2.2 POSITIVE AND NEGATIVE SYNDROME SCALE (PANSS)

The Positive and Negative Syndrome Scale (PANSS) is a medical scale used for measuring symptom severity of patients with schizophrenia. The patient is rated from 1 to 7 on 30 different symptoms based on the interview as well as reports of primary care hospital workers. It has 3 catalogs, positive scale (p1-p7 in our research), negative scale (m1-m7 in our research) and general

psychopathology scale (g1-g16 in our research). The minimum score of PANSS total is 30, and the maximum score of PANSS is 210. It should be noted that 1 rather 0 is the lowest score for each item [4].

According to definition of PANSS, just SZs (90) did this scale. Except the original 30 items, our cooperating parties combined some items together as new items based on their features (see ../../SZData/panss.txt).

3 Methods

According to MCCB and PANSS scale, using R package (glmnet) to do regression. The elasticnet mixing parameter, alpha, with $0 \leqslant \alpha \leqslant 1$. alpha=1 is the lasso penalty, and alpha=0 the ridge penalty.
The penalty is defined as

$$(1- \alpha)/2||\beta||\_2^2+\alpha||\beta||\_1.$$

Lambda was choose by using different cross validation ways. In this report, firstly I used training set : testing set (3:1), 5 fold CV and then I used Leave One Out Cross Validation (LOOCV). Specific steps is below.
To tell if the model fits the data, we use pseudo R-squared, defined as:

$$R^2_{=1} - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Where N is the numbers of observations in the model, is the i observation of the clinical indices and is it predicted value of , while represents the mean of the clinical indices.

3.1 training and testing

As for PANSS, seed 1, gaussian family, data set was selected randomly with training set (67 samples) and testing set (23 samples). No phenotype's pseudo R2 is larger than 0.15.
As for MCCB, gaussian family and poisson family, total 126 samples had MCCB scale, data set was selected randomly with training set (94 samples : 40 HCs and 54 SZs) and testing set (32 samples : 14 HCs and 18 SZs). No phenotype's pseudo R2 is larger than 0.15.

3.2 LOOCV

With LOOCV, although training set is larger than above methods, there still no MCCB's pseudo R2 is larger than 0.15 with Gaussian or poisson family. But in some phenotypes in PANSS scale, this

model can fit it well. I choose the phenotype with pseudo R2 > 0.15, and write down the markers of fitted model which named by "phenotype name" using guassian family, and "phenotype name poisson" using poisson family.