

Curve

Introduction

To identify the relationship of mOTU and PANSS, curve figures were plotted.

Data information

360 mOTUs profile which represents bacteria abundance was calculated by analyzing the sequencing data. As the metagenomics data, mOTUs profile contains numerous zero. Each column of this mOTUs profile is the total bacteria and it's abundance in this sample. Each row of this mOTUs profile is the abundance of this bacteria in total samples. In other words, this profile represents 360 mOTUs' (row) abundance in 171 samples.

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. According to definition of PANSS, just SZs (90) did this scale.

Except the original 30 items (Figure 1), our cooperating parties combined some items together as new items based on their features (../SZData/panss.txt), and this new type with 27 items (Figure 2) is we usually used, such as positive factor 1.

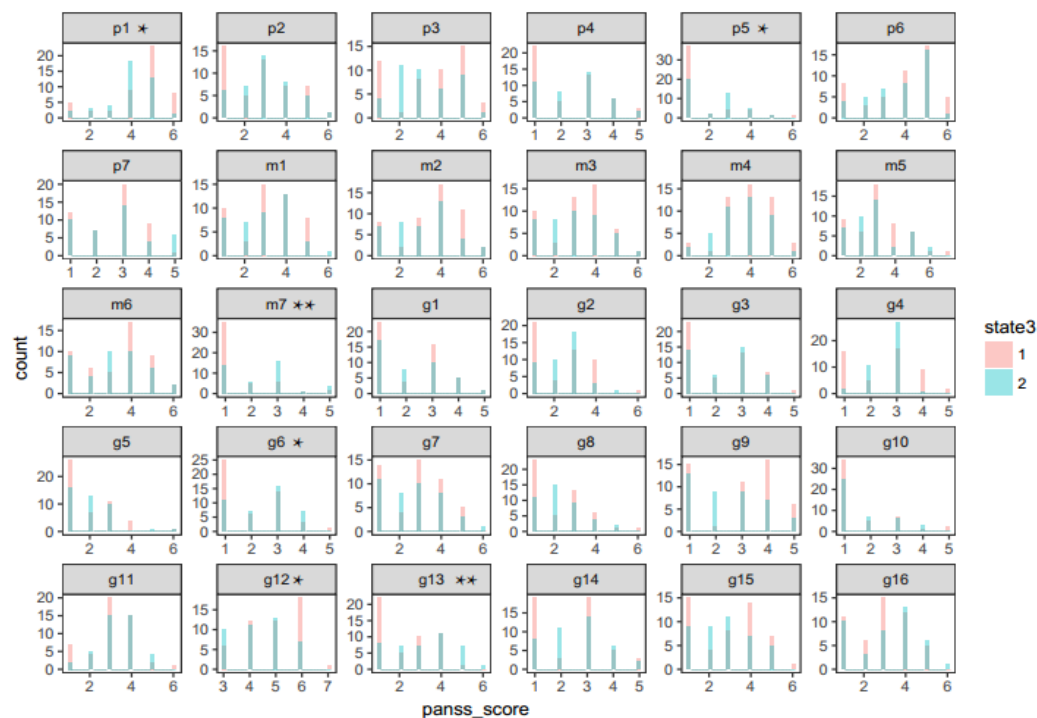


Figure 1. Histogram of the original 30 items of PANSS scale.

The data histogram of each items between first episode (n = 49, red) compared with relapse (n = 41, blue). And the x axis is the score of each sub figure.

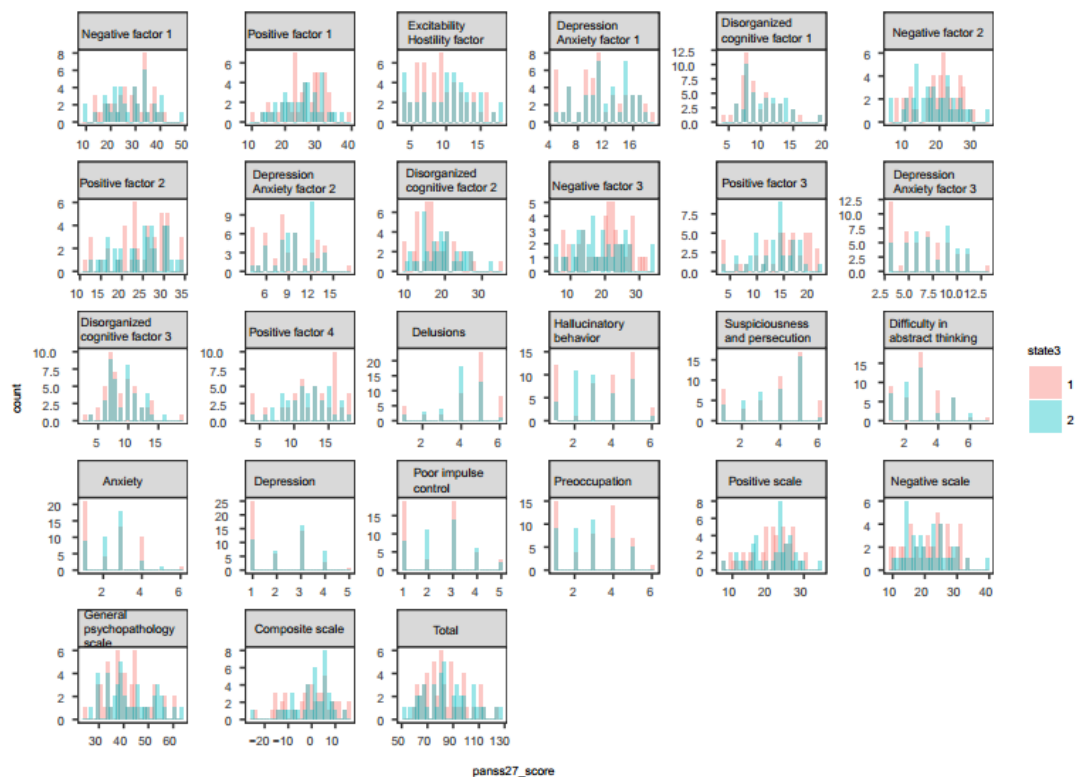


Figure 2. Histogram of the new type with 27 items of PANSS scale.

The data histogram of each items between first episode (n = 49, red) compared with relapse (n = 41, blue). And the x axis is the score of each sub figure.

What's more, for pdf version of these two figures, please look up the same directory. Figure 1 is named by panss_trem30_0915.pdf and Figure 2 is named by panss_term27_1117.pdf

Procedure

To identify the relationship of mOTU and PANSS, it's better to show is this way each curve represents the relationship of one mOTU abundance (Y) and matched one PANSS score (X). But X must be unique with just one mOTU abundance. So in this situation, I mean these two sample's mOTUs respectively. In the end, the X is unique with matching Y. To show it clearly, I made a example below with R code (Figure 3 a,b).

```
> x <- c(10, 10, 11, 12, 12, 13)      # one of PANSS score
> y <- c(0.001, 0.2, 0.001, 0.02, 0.005, 0.000005)  # one of mOTU profile
```

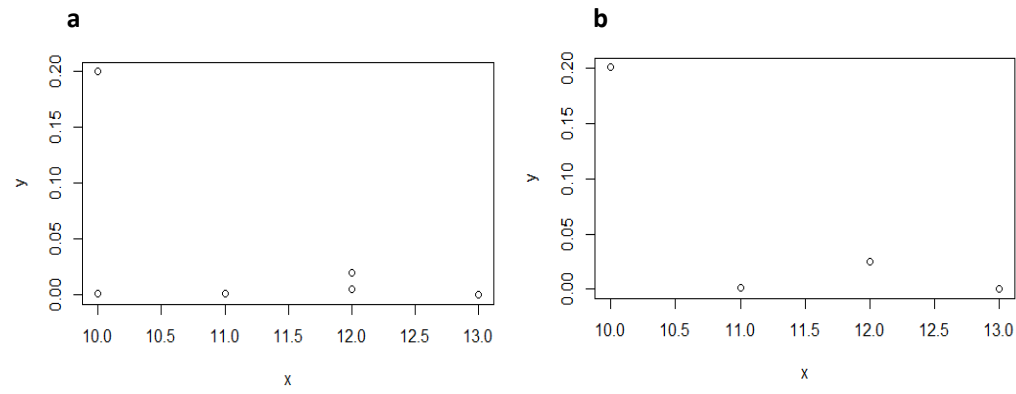


Figure 3 Example of scatter plot with PANSS score (x) and mOTU profile(y).

(a) The original x and y. In some cases, the x has two y values. (b) process data. If the original x has two y values, then mean the y, and unique x.

Although I can draw the total PANSS scales with 27 items, I just push two html figure on Github to give a glance. The thing I wonder is if we should preprocess the data, such as $\log(x)$?