

# Real data analysis outlines

*February, 2019*

The dataset contains 37 kinds of microbiomes and 499 samples at different times: 3, 6, 10, 12 weeks separately.

We would like to illustrate our convex bi-clustering methods in terms of this real data, i.e. whether the method can identify different treatment groups ('Control', 'PAT') and whether the method can separate different groups of microbiomes. However, here, although we have known the labels of the two intervention arms, we do not have more information about the microbiomes. Therefore, we may increase and decrease some values of the selected microbiomes to group them by ourselves at the very beginning. The convex bi-cluster methods are applied later to identify different treatment groups and different microbiome groups, which were made by ourselves.

Here is the outline of the real data analysis:

## **A. Pure real data (without any change on the real data)**

- 1. First, we illustrated whether there are differences between the treatment group: PAT and the control group. A two-way ANOVA was applied for that propose.
- 2. Second, we test whether our clustering method can separate those two groups

## **B. Semi-real data on the subjects.**

Highlight: Select one group, e.g. the control group. First, test whether the distributions of microbiomes are different within the group. If not, select half of the subjects as group 1 and the rest as group 2. Increase or decrease some microbiome values in group 1. Then apply our convex bi-clustering method to test whether we can separate group 1 and group 2.

### **More details:**

- 1. Within one intervention group, e.g. control group, we randomly selected two subjects to compare their microbiome counts through ANOVA to see whether there are differences within the group.
- 2. Sort the columns of the dataset. Make the values in each column have a similar order.
- 3. Divide the subjects into two groups: group 1, group 2.
- 4. Multiple the part of the microbiomes in group 1 with an arbitrary value larger than 1.
- 5. Multiple the part of the microbiomes in group 2 with an arbitrary value less than 1.
- 6. Apply ANOVA again to check whether those two groups have differences and check whether the microbiomes within one group have differences.
- 7. Apply the convex bi-clustering methods to check whether they can identify different groups of subjects and different groups of microbiomes.

## **C. Semi-real data on both subjects and microbiomes.**

We would like to see whether our method can this help us separate microbiome groups when we have already known the treatment groups.

(I would like to check this is due to the results in B are not very ideal. When we can have a good division in microbiomes, the clustering in subjects was not very well)

**Steps:**

- 1. Combine subjects and microbiomes from the PAT group and control group.
- 2. Fill NAs with 0s.
- 3. Sort the columns of the dataset. Make the values in each column have a similar order.
- 4. Divide the subjects into two groups: group 1, group 2.
- 5. Multiple the part of the microbiomes in group 1 with an arbitrary value larger than 1.
- 6. Multiple the part of the microbiomes in group 2 with an arbitrary value less than 1.
- 7. Apply the convex bi-clustering methods to check whether they can identify different groups of subjects and different groups of microbiomes.