Assignment

You have recently sequenced the gut microbiomes from 82 Covid-19 patients with varying degrees of Covid Severity (determined based on the WHO Severity standard). You have the species-level centered log-ratio (clr) transformed abundance profiles of the various taxa in the gut microbiome (provided in the file Assignment\_ClrTrans\_Species.xlsx). You also have the metadata of these patients which contains information pertaining to the Severity of the Covid-19 (ordinally arranged into three different groups in increasing order of severity namely, mild, moderate and critical\_severe (Assignment\_Metadata.xlsx). You can convert these three numeric encodings of 1, 2 and 3).

Besides the Covid severity, you also have the metadata pertaining to the age, gender and BMI of the patients. In addition, you also have information pertaining to 12 different co-morbidities: Hypertension (abbreviated as HTN), Diabetes, Respiratory Disease, Heart Disease, Renal Disease, Liver Disease, Malignancy, Immunosuppressive Disease, Obesity, Neurological Disease, Metabolic Disease and Cardiovascular Disease. The total number of co- morbidities per patient is also indicated.

You have to investigate the following:

1. a. Is there a significant variation in the Evenness (Pielou Evenness) and overall alpha diversity (both Richness and Evenness quantified by the Shannon Index) between the patients across the different severity groups? Show the results as boxplots and indicate the appropriate tests used.

b. Is there a difference in the overall abundance pattern of the microbiomes (beta-diversity) across the different severity groups. Show the results as Prinicipal Component Analysis or Principal Coordinate Analysis plots and indicate the significance of separation. Which test would you use for this?

1. a. Which species show a significant (Benjamini-Hochberg corrected FDR

<= 0.1) association with Covid-19 severity without adjusting for any confounders? Which species show a significant (Benjamini-Hochberg corrected FDR <= 0.1) association with Covid-19 severity after adjusting for age, BMI, gender and total number of comorbidities? Take only those that are detected in at least 5% of the samples.

1. Take all metadata columns, Age, Gender, BMI, WHO Severity, the total number of comorbidities and the metadata pertaining to the presence of each of the 12 co-morbidities. Compute the association of the overall gut microbiome composition with each of these metadata. Which tests do we use for that? Arrange the metadata in decreasing order of association with the overall gut microbiome composition. Which metadata show significant association (P <= 0.05)
2. Investigate the possibility of predicting WHO-based Covid-19 Severity using the species-level gut microbiome profile and specific metadata: Age, BMI, Gender and the total number of comorbidities. Compare the relative classification accuracy obtained using the three machine-learning based classifiers: Random Forest, Support Vector Machines and k-Nearest Neighbors. Identify the top 50 predictive features for this scheme.

First investigate the classification performance on the three different groups. Then investigate the classification performance only between mild and critical\_severe groups.