HMDS_assignment_Group_2

2023-03-18

R Markdown

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.

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.6-4
```

```
library(ecodist)
```

```
##
## Attaching package: 'ecodist'
```

```
## The following object is masked from 'package:vegan':
##
##     mantel
```

```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("ggplot2")
```

##Reading the dataset files

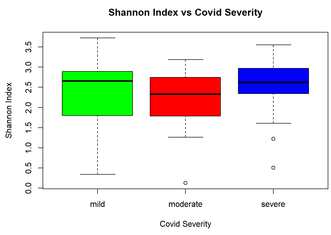
```
data1_rawcount <- read.csv("Assignment_RawCount_Species - Sheet1.csv")
data2_metadata<-read.csv("Assignment1_Metadata - Sheet1.csv")
```

```
c<-data2_metadata$X #Retrieving first column from metadata file
df_merged <- data1_rawcount[data1_rawcount$X %in% c,] #Rows from rawcount are copied based on column X in rawcoun
t and metadata files
```

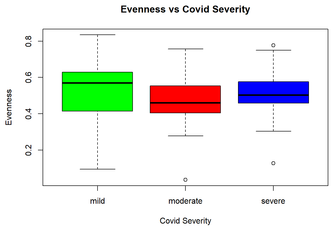
```
#Calculating Shannon Index
df_shannon=(diversity(df_merged[-1], index="shannon"))
df_merged['Shannon Index']=df_shannon                  #Adding Shannon Index values to df_merged


#Adding WHO_severity column from metadata to df_merged
df_merged$WHO_severity <- data2_metadata$WHO_severity[match(df_merged$X, data2_metadata$X)]

#Converting mild,moderate,critical_severe to 1,2 and 3
df_merged$WHO_severity[which(df_merged$WHO_severity=="mild")] <- 1
df_merged$WHO_severity[which(df_merged$WHO_severity=="moderate")]<-2
df_merged$WHO_severity[which(df_merged$WHO_severity=="critical_severe")] <- 3
```

```
#Calculating Evenness
S <- apply(df_merged[,-1]>0,1,sum)
df_evenness<-df_shannon/log(S)
df_merged['Evenness']<-df_evenness #Adding evenness column to df_merged
```

```
#Plotting Shannon Index vs Covid Severity
x<- df_merged$WHO_severity
y<-df_merged$`Shannon Index`
boxplot(y~ x, data = df_merged,names=c("mild", "moderate","severe"),xlab = "Covid Severity",ylab = "Shannon Inde
x",main = "Shannon Index vs Covid Severity",col = c("green", "red", "blue"))
```




```
#Plotting Evenness vs Covid Severity
x<- df_merged$WHO_severity
y1<-df_merged$Evenness
boxplot(y1~ x, data = df_merged,names=c("mild", "moderate","severe"),xlab = "Covid Severity",ylab = "Evenness",ma
in = "Evenness vs Covid Severity",col=c("green","red","blue"))
```



After visualizing the boxplots, we have performed statistical tests to determine if there is a significant difference in the alpha diversity and evenness measures between the Covid severity groups.Since the data is not following a normal distribution, we can also see outliers in box plot to give this statement a prove Thus we can use non-parametric tests such as the Kruskal-Wallis test to compare the distributions of alpha diversity and evenness measures across the different Covid severity groups.We can perform these tests using the stats package in R. The code for performing the tests is as follows:

Null hypothesis : there is no significant variation in the evenness or overall alpha diversity Alternate: there is significant variation in the evenness or overall alpha diversity

```
# Kruskal-Wallis test for Shannon Index - overall alpha diversity
kw_diversity <- kruskal.test(y ~ x, data = df_merged)
print(kw_diversity)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  y by x
## Kruskal-Wallis chi-squared = 2.7838, df = 2, p-value = 0.2486
```

```
# Kruskal-Wallis test for Pielou Evenness
kw_evenness <- kruskal.test(y1 ~ x, data = df_merged)
print(kw_evenness)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  y1 by x
## Kruskal-Wallis chi-squared = 3.5061, df = 2, p-value = 0.1732
```

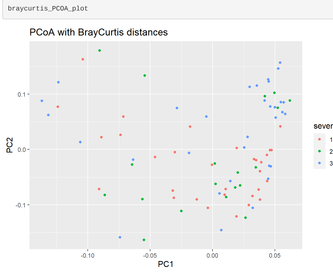
Based on the Kruskal-Wallis rank sum test results, there is no significant variation in the evenness or overall alpha diversity (both richness and evenness quantified by the Shannon index) between the patients across the different severity groups. The p-values for both tests are greater than 0.05(kw_diversity=0.2486 and kw_evenness=0.1732), indicating that we cannot reject the null hypothesis that there is no significant difference between the groups.

1. 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?

```
##Creating a dataframe to calculate beta diversity
data2_metadata<-read.csv("Assignment1_Metadata - Sheet1.csv")
data3_clrcount <- read.csv("Assignment1_ClrTrans_Species - Sheet1.csv")


df_merged_clr <- data3_clrcount[data3_clrcount$X %in% data2_metadata$X,]
mat_df.mdf <- as.matrix.data.frame(df_merged_clr[-1])
```

```
###Calculating the distance matrix
div_beta=vegdist(mat_df.mdf,method="bray")
res<-pco(div_beta)

#Adding WHO_severity column from metadata to df_merged
df_merged_clr$WHO_severity <- data2_metadata$WHO_severity[match(df_merged_clr$X, data2_metadata$X)]

#Converting mild,moderate,critical_severe to 1,2 and 3
df_merged_clr$WHO_severity[which(df_merged_clr$WHO_severity=="mild")] <- 1
df_merged_clr$WHO_severity[which(df_merged_clr$WHO_severity=="moderate")]<-2
df_merged_clr$WHO_severity[which(df_merged_clr$WHO_severity=="critical_severe")] <- 3
###Adding PC1,PC2 and WHO Severity to res_pcoa_df
res_pcoa_df <- data.frame(pcoa1 = res$vectors[,1], pcoa2 = res$vectors[,2])

res_pcoa_df <- cbind(res_pcoa_df,severity=df_merged_clr$WHO_severity)
```

```
###Plotting PCOA using data=res_pcoa_df

braycurtis_PCOA_plot <- ggplot(data = res_pcoa_df, aes(x=pcoa1, y=pcoa2,
                                                       color = severity)) + geom_point() +labs(x = "PC1",y = "PC
2",title = "PCoA with BrayCurtis distances") +theme(title = element_text(size = 12))

braycurtis_PCOA_plot
```



This will produce a PCoA plot of the gut microbiome beta-diversity, with the points colored by the severity of Covid-19. We can use PERMANOVA (permutational multivariate analysis of variance) to test whether there is a significant difference in the microbiome composition across the different severity groups:

```
pcoa_adonis <- adonis2(div_beta ~ WHO_severity, data = data2_metadata, permutations = 999)
print(pcoa_adonis)
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = div_beta ~ WHO_severity, data = data2_metadata, permutations = 999)
##              Df SumOfSqs      R2      F Pr(>F)
## WHO_severity  2   0.6574 0.03659 1.5001  0.049 *
## Residual     79  17.3108 0.96341
## Total        81  17.9682 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of 0.045 suggests that there is a significant difference in the gut microbiome composition across the different severity groups of COVID-19 in dataset.