Statistical methods for microbiome análisis

Practical exercises

Data set Qin et al. (2012) Type2 Diabetes microbiome study: <https://www.nature.com/articles/nature11450>

1. Upload the data as a phyloseq object as follows:

data<-readRDS("qin\_all.rds")

1. Use phyloseq library to preprocess and describe the dataset:
   1. Number of taxa, number of samples, number of variables in metadata
   2. Agglomerate at the genus level and normalize the data to relative abundances
   3. Summarize the data at the genus level: number of taxa and summary and histogram of log10(mean taxa abundances)
   4. Filter out those taxa at the genus level with a mean relative abundance larger than 0.001
   5. Normalize the data at genus level to relative abundances
   6. Compute alpha diversity Shannon index and the effective number of taxa as follows:

rich<-estimate\_richness(data, measures =c("Shannon"))

effnum <- exp(rich$Shannon)

* 1. Test for differences in the effective number of species between cases and controls (boxplot and Wilcoxon test)
  2. Perform ordination plots using both MDS and NMDS and the BC, UF and wUF distances. Display the samples according to disease.
  3. Test for global differences between cases and controls with PERMANOVA

1. Use coda4microbiome R package to analyze qin dataset at genus level
   1. Perform an exploratory analysis of log-ratios
   2. Implement variable selection to obtain a microbial signature that is predictive of disease status