USER MANUAL FOR ANCOM 2.0

The current code implements ANCOM in cross-sectional and longitudinal datasets while allowing the use of covariates. The following libraries need to be included for the R code to run:

library(exactRankTests)

library(nlme)

library(dplyr)

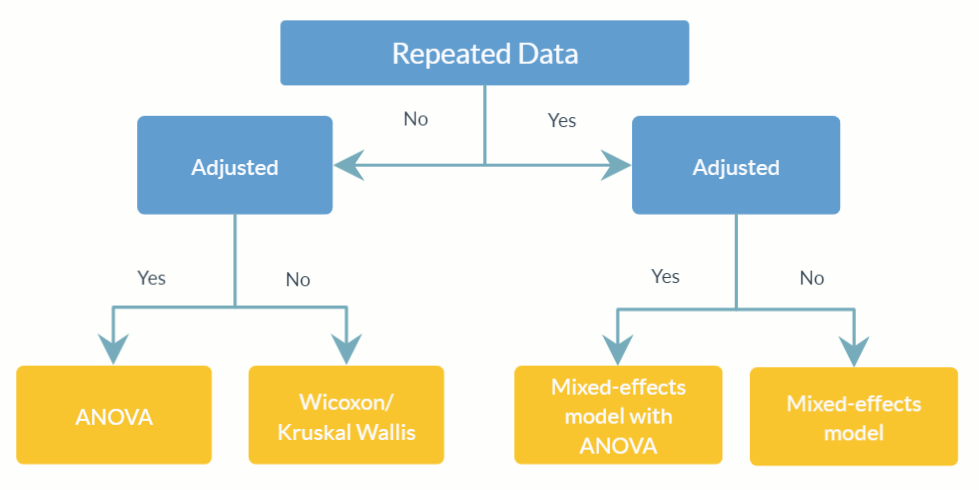
# Input requirements

* OTU data or taxa data: This is should be a matrix or data frame with each sample in rows and OTUs in columns. OTU data should contains sample identifier with column name “Sample.ID”.
* Metadata: This is the datafile with all variables and covariates of interest. It should be a matrix or data frame containing the sample identifier with column name “Sample.ID” and each following column being the variables.

# Arguments of the function (ANCOM)

* otu\_data: Data frame representing OTU data or taxa data (must meet specifications as mentioned earlier).
* meta\_data: Data frame of variables (must meet specifications as mentioned earlier).
* main\_var: Character. The name of the main variable of interest.
* zero\_cut: Numeric fraction. OTUs with proportion of zeroes greater than zero\_cut are not included in the analysis. Default is 0.90.
* p\_adjust\_method: Character. Specifying the method to adjust p-values for multiple comparisons. Default is “BH” (Benjamini-Hochberg procedure).
* alpha: Level of significance. Default is 0.05.
* adj\_formula: Character string representing the formula for adjustment (see example).
* rand\_formula: Character string representing the formula for random effects in lme (see example).

# A flowchart of the tests within ANCOM



# Function outputs

* res: A data frame with the W statistic for each taxa and subsequent columns which are logical indicators of whether an OTU or taxon is differentially abundant under a series of cutoffs (0.9, 0.8, 0.7 and 0.6).

# Examples

## 1. Standard analysis

## 

### *# Detection of differentially abundant OTU between subjects*

### *# Example dataset: moving-pics*

res = ANCOM(otu\_data = otu\_data,

meta\_data = meta\_data,

main\_var = “Subject”,

zero\_cut = 0.90,

p\_adjust\_method = “BH”,

alpha = 0.05,

adj\_formula = NULL,

rand\_formula = NULL)

## 2. Adjusted for covariates

### *# Detection of differentially abundant OTU between subjects adjusted for antibiotic usage*

### *# Example dataset: moving-pics*

res = ANCOM(otu\_data = otu\_data,

meta\_data = meta\_data,

main\_var = “Subject”,

zero\_cut = 0.90,

p\_adjust\_method = “BH”,

alpha = 0.05,

adj\_formula = ”ReportedAntibioticUsage”,

rand\_formula = NULL)

## 3. Repeated measure/longitudinal analysis

### 3.1 Random intercept model

### *# Detection of differentially abundant genera between delivery methods accounting for random subject effect*

### *# Each subject has his/her own intercept*

### *# Example dataset: ecam*

res = ANCOM(otu\_data = otu\_data,

meta\_data = meta\_data,

main\_var = "delivery",

zero\_cut = 0.90,

p\_adjust\_method = “BH”,

alpha = 0.05,

adj\_formula = NULL,

rand\_formula = "~ 1 | studyid")

### 3.2 Random intercept model adjusted for other covariates

### *# Detection of differentially abundant genera between delivery methods accounting for fixed time effect and random subject effect*

### *# Each subject has his/her own intercept*

### *# Example dataset: ecam*

res = ANCOM(otu\_data = otu\_data,

meta\_data = meta\_data,

main\_var = "delivery",

zero\_cut = 0.90,

p\_adjust\_method = “BH”,

alpha = 0.05,

adj\_formula = ”month”,

rand\_formula = "~ 1 | studyid")

### 3.3 Random coefficients/slope model

### *# Detection of differentially abundant genera between delivery methods accounting for random time effect and random subject effect*

### *# Each subject has his/her own intercept and slope*

### *# Example dataset: ecam*

res = ANCOM(otu\_data = otu\_data,

meta\_data = meta\_data,

main\_var = "delivery",

zero\_cut = 0.90,

p\_adjust\_method = “BH”,

alpha = 0.05,

adj\_formula = ”month”,

rand\_formula = "~ month | studyid")